

# eRevise+RF: A Writing Evaluation System for Assessing Student Essay Revisions and Providing Formative Feedback

Zhexiong Liu<sup>1</sup>, Diane Litman<sup>1,2</sup>, Elaine Wang<sup>3</sup>, Tianwen Li<sup>2</sup>, Mason Gobat<sup>1</sup>,  
Lindsay Clare Matsumura<sup>2</sup>, Richard Correnti<sup>2</sup>

<sup>1</sup>Department of Computer Science, <sup>2</sup>Learning Research and Development Center  
University of Pittsburgh, RAND Corporation, Pittsburgh PA USA  
zhexiong@cs.pitt.edu, ewang@rand.org  
{dlitman,mng28,tianwen.li,lclare,rcorrent}@pitt.edu

## Abstract

The ability to revise essays in response to feedback is important for students' writing success. An automated writing evaluation (AWE) system that supports students in revising their essays is thus essential. We present eRevise+RF, an enhanced AWE system for assessing student essay revisions (e.g., changes made to an essay to improve its quality in response to essay feedback) and providing revision feedback. We deployed the system with 6 teachers and 406 students across 3 schools in Pennsylvania and Louisiana. The results confirmed its effectiveness in (1) assessing student essays in terms of evidence usage, (2) extracting evidence and reasoning revisions across essays, and (3) determining revision success in responding to feedback. The evaluation also suggested eRevise+RF is a helpful system for young students to improve their argumentative writing skills through revision and formative feedback.

## 1 Introduction

Young student writers often struggle with identifying convincing evidence and connecting it to claims when writing argumentative essays; however, developing persuasive argumentation itself helps students improve their thinking and reasoning skills (Kuhn et al., 2017). This motivates recent research in developing Automated Writing Evaluation (AWE) systems for supporting students in writing and revising argumentative essays. For instance, Zhang et al. (2019) developed the eRevise system to score student essays on text-based evidence usage and to provide associated feedback to guide revision. Although students attempted to respond to the feedback, their revisions often did not yield substantive essay improvement (Wang et al., 2020). Similar findings have been found for other AWE systems (Graham et al., 2015), suggesting that students often lack the skills necessary for effective revision (Roscoe et al., 2013).

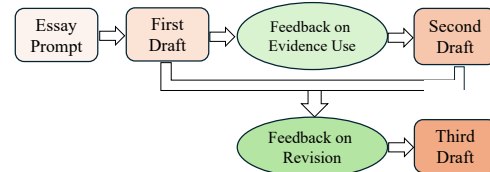


Figure 1: The eRevise+RF system usage pipeline, where students work on three essay drafts, and receive evidence use feedback and revision feedback.

The feedback provided by Zhang et al. (2019) was limited to evidence use, i.e., “Adding more evidence would make your argument even more convincing.” Additional feedback on revisions made in response to argumentative feedback could potentially help students self-evaluate whether their revision attempts indeed improve their essays. For instance, the feedback “When you revised your essay, it looks like you added in evidence that was very similar to the evidence you had included before. When writers revise, they generally add new content to their essays” suggests an attempt to address feedback on evidence use was unsuccessful, because existing rather than new evidence was added. We envision that using NLP to provide feedback addressing revision will further support the development of students’ argumentative writing skills.

Motivated by this, we developed the eRevise+RF system<sup>1</sup> (RF stands for “revision feedback”) to assess whether student revisions align with the previous feedback they receive and examine how these revisions successfully improve the argumentative essay. In particular, our system is used to support students who are taking the Response-to-Text Assessment (RTA) (Correnti et al., 2013), which is an argumentative writing task for assessing students’ ability to reason using text-based evidence and to successfully use evidence to support their claims. Research on the technical quality of RTA

<sup>1</sup>The eRevise+RF system is online at <http://erevise.lrdc.pitt.edu>. The source code is available at <https://github.com/ZhexiongLiu/eRevise-RF-System>

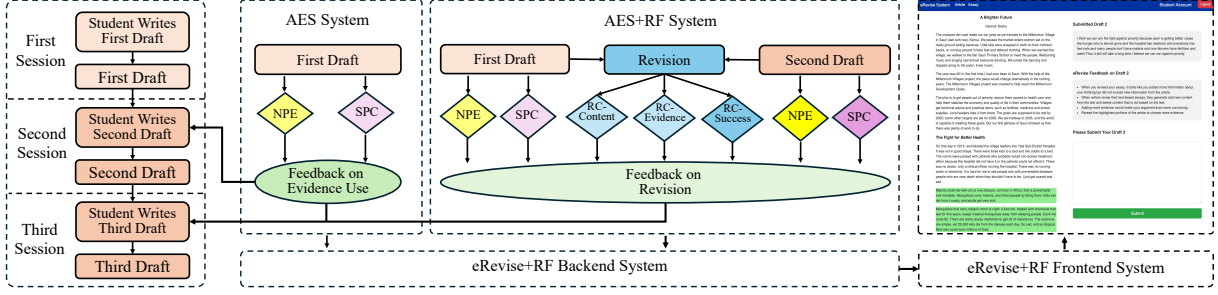


Figure 2: System architecture, including AES and AES+RF backend systems and user frontend interface. The AES system generates evidence use feedback based on scoring indicators (NPE and SPC scores) and the AES+RF system provides revision feedback based on both scoring indicators and revision classifiers.

scores indicates strong evidence of construct validity as assessed by examining the relationship of scores to measures of literacy instruction and student achievement (Correnti et al., 2013, 2020). To administer the RTA, a teacher reads a non-fiction article aloud to students as they follow along with their copy. For our pilot studies, students read an article about the United Nations Millennium Villages Project to fight poverty in Kenya (MVP) in Appendix A, or an article about the benefits and costs of Space Exploration (SPACE) in Appendix B.

Figure 1 shows the system usage pipeline after reading the RTA article. First, students use the system to write an essay in response to a prompt, e.g., “Based on the article, did the author provide a convincing argument that winning the fight against poverty is achievable in our lifetime? Explain why or why not with 3 to 4 examples from the text to support your answer.” After students submit their first essay drafts, the backend of the system automatically processes the essays by extracting evidence features (e.g., scoring indicators), which are then used to select from expert-designed messages to provide feedback on text-based evidence usage. After students receive this feedback and revise and submit second drafts of their essays, the backend system now extracts all revised sentences between the first and the second drafts and classifies whether each revision contributes to an essay’s improvement in alignment with the system’s prior feedback. Both the evidence features and a set of revision classifiers are used to select from expert-designed messages to provide feedback on revisions and guide students in writing their third drafts. Here, we use expert-crafted rather than LLM-generated feedback because expert feedback offers more specific and actionable suggestions to address critical issues in student essays, which is particularly helpful for young student writers. Similar findings have been confirmed in a recent study (Behzad et al., 2024).

We evaluate eRevise+RF by addressing two research questions: **RQ1**: Can the system use NLP to effectively assess student argumentative writing and revisions? **RQ2**: Can the system help students improve essays through formative feedback on their revision quality and evidence usage?

## 2 Related AWE Work on Revision

While AWE systems typically allow submissions of multiple drafts in response to feedback, most provide feedback on single essay drafts rather than on the revisions between two drafts (Wilson et al., 2021; Huawei and Aryadoust, 2023; Fleckenstein et al., 2023). However, feedback on revisions has been identified as an important area for AWE research (Guo et al., 2023; Correnti et al., 2024; Li et al., 2024), and some systems have visualized NLP revision analyses, e.g., to assist students in self-monitoring writing (Zhang et al., 2016; Shibani et al., 2018; Litman et al., 2022). While these systems display revisions, they do not provide feedback messages based on an NLP assessment of a revision’s success. Although prior NLP revision work has created annotated revision corpora (Zhang et al., 2017; Anthonio et al., 2020; Spangher et al., 2022; Du et al., 2022; D’Arcy et al., 2024) and has developed models for isolated tasks like identifying revision purposes (Afrin et al., 2020; Kashefi et al., 2022; Du et al., 2022; Mita et al., 2024; Jourdan et al., 2024), generating revisions (Chong et al., 2023; Ziegenbein et al., 2024), or evaluating revision quality (Liu et al., 2023), our work is the first to integrate multiple such tasks into a deployed and evaluated AWE system.

## 3 System Architecture

### 3.1 eRevise+RF Backend

The backend system (in Figure 2) builds on and integrates NLP algorithms from our prior re-

search (Zhang et al., 2019; Liu et al., 2023) to process argumentative essays, revisions between drafts, and provide feedback on both. It consists of an AES (automated essay scoring) system and an AES+RF system, of which the former provides evidence use feedback and the latter focuses on revision feedback. The AES system uses NLP algorithms to extract a set of scoring indicators and then uses the indicators to select from a set of expert-designed evidence feedback messages. The AES+RF system uses revision classifiers in addition to scoring indicators to select from another set of expert-designed revision feedback messages.

### 3.1.1 Evidence Scoring Indicators

The algorithms used in the AES and AES+RF systems to extract features from essays were adapted from the eRevise system (Zhang et al., 2019). Given the text of an essay draft, the algorithm computes the Number of Pieces of Evidence (NPE), an integer encoding the number of evidence topics mentioned in the essay<sup>2</sup>. A sliding window is applied to extract NPE on the raw text. If a window contains a similar word from a manually crafted list of keywords that are associated with evidence topics, the window is confirmed to contain text-based evidence that is related to the topic. The eight-word sliding window, Glove embedding (Pennington et al., 2014), cosine similarity with a 0.9 threshold are used in the implementation based on the training data from previous RTA deployments (Zhang et al., 2019). Also, the indicator Specificity (SPC) of the essay uses the same sliding window to determine if it contains words from another manually crafted list, which is a vector of integers that encodes the number of specific examples mentioned in 8 categories for MVP and 7 categories for SPACE, taken from Rahimi et al. (2017). SPC score is the sum of the vector, whose value is a positive integer. The NPE and SPC compute the breadth and depth of the text-based evidence usage in an essay, respectively.

### 3.1.2 Revision Classifiers

In addition to the scoring indicators, the AES+RF system uses classifiers to determine if a sentence-level revision successfully contributes to an essay improvement in alignment with the feedback received. Specifically, the original and revised

drafts are aligned into pairs of original and revised sentences using the sentence alignment tool Bertalign (Liu and Zhu, 2022). The pairs of non-identically aligned sentences are extracted as the essay revisions. A BERT-based (Devlin et al., 2019) binary revision purpose classifier (**RC-Content**) that is trained on a college-level revision corpus (Kashefi et al., 2022) is used to classify essay revisions into surface (meaning-preserving) and content (meaning-altering) revisions. Note that the surface revisions are later removed because the system is only interested in content revisions. Another binary revision classifier (**RC-Evidence**) is used to predict whether each content revision is an evidence or reasoning revision, motivated by the prior work that assesses revision quality from the perspective of evidence and reasoning (Afrin et al., 2020). RC-Evidence is implemented with ChatGPT, which returns an evidence or reasoning label given a revision pair and the prompt “*You need to identify whether the given sentence is an evidence or reasoning sentence. Your output should be chosen from the list [evidence, reasoning]*”. Furthermore, a binary revision classifier (**RC-Success**) predicts whether each content revision is successful or unsuccessful. RC-Success is a DistilRoBERTa model (Sanh et al., 2019) developed based on Liu et al. (2023). It leverages Argument Context (AC) in addition to revision pairs to determine successful vs. unsuccessful revisions. The ACs are extracted from essays with Chain-of-Thought prompts for ChatGPT used in prior work (Liu et al., 2023). Finally, the combination of RC-Evidence and RC-Success yields 4 revision labels, i.e., successful evidence, unsuccessful evidence, successful reasoning, and unsuccessful reasoning.

### 3.1.3 Evidence Use Feedback

The AES system provides 3 levels of Evidence Use Feedback (EF). EF1 feedback focuses solely on providing more evidence (completeness); EF2 feedback focuses on providing more details (specificity); and EF3 feedback focuses on explaining and connecting evidence to an argument (explanation). The AES system uses scoring indicators (NPE and SPC) with expert-designed thresholds ( $\alpha$  and  $\beta$ ) to determine which feedback level message to display after essay submission. An essay will receive EF1 if its NPE score is less or equal to  $\alpha$ , which indicates no more than half of the topics are mentioned by the essay. In this case, we highlight text portions with a green background

<sup>2</sup>The expert-crafted evidence topics include Hospital, Malaria, Farming, and School for MVP article; People, Earth, Cost, and Exploration for SPACE article (Rahimi et al., 2017).

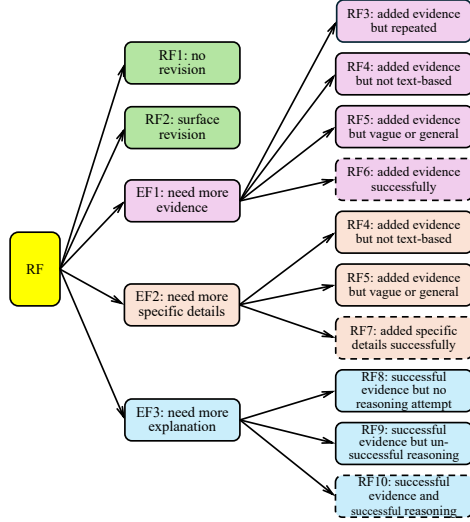


Figure 3: Revision feedback tree, where solid squares are unsuccessful revision feedback focused on helping the student try again, and the dotted squares are successful revision feedback that advances the student to a new evidence usage skill.

in the frontend (see Figure 5 in the Appendix) to help students focus on missing topics. An essay will receive EF2 if the NPE score is greater than  $\alpha$  and SPC is less than or equal to  $\beta$ , which indicates the essay discusses several topics but details are still missing. An essay will receive EF3 if the NPE score is greater than  $\alpha$  and SPC is greater than  $\beta$ , which indicates that since the completeness and specificity of evidence are satisfactory, the student should focus on explaining how each piece of evidence is connected to a claim. The threshold  $\alpha$  for NPE is set to half of the topics discussed in the relevant RTA source article, and  $\beta$  for SPC is set to half of the example categories. Specifically,  $\alpha$  and  $\beta$  are set to 2 and 4. The feedback messages and how they are selected are shown in Table 5 in the Appendix.

### 3.1.4 Revision Feedback

The AES+RF system provides 10 levels of Revision Feedback (RF) organized on a revision feedback tree in Figure 3, after assessing whether a student successfully follows the EF to revise the last essay draft. Thus RF focuses on the revision process rather than on the essay content itself. For a student who attempts to revise based on EF previously received, either unsuccessful revision feedback is provided to help the student try again, or successful revision feedback is used to help advance the student’s essay into the next level. The system uses scoring indicators and revision classifiers to select RF from the revision feedback tree.

Specifically, *Path RF-RF1*: In the case a student makes no revision when all aligned revision pairs are identical, the feedback is RF1. *Path RF-RF2*: If a student makes only surface revisions as predicted by RC-Content, RF2 is used. In the cases where the last draft had EF1, the RF has 4 sub-paths. *Path RF-EF1-RF3*: RF3 is given if there are no NPE changes between the drafts, and the newly added sentences contain topic words that already exist. *Path RF-EF1-RF4*: RF4 is used if NPE has no increases between drafts, and added sentences contain no topic words. *Path RF-EF1-RF5*: It provides RF5 if NPE increases, but the increment of SPC is less than an expert-designed threshold  $\gamma$  ( $\gamma$  is set to 2). *Path RF-EF1-RF6*: RF6 recognizes the student’s successful revisions in response to EF1 (completeness), thus needs to focus on EF2 (specificity). In the cases where EF on the last draft had EF2, the RF has 3 sub-paths. *Path RF-EF2-RF4*: RF4 is used if SPC increment is less than  $\gamma$ , and added sentences contain no SPC words<sup>3</sup>. *Path RF-EF2-RF5*: RF5 is used if SPC increment is less than  $\gamma$  but contains SPC words. *Path RF-EF2-RF7*: RF7 is selected when the revision successfully addressed EF2. In the cases where EF on the last draft was EF3, RF has 3 sub-paths. *Path RF-EF3-RF8*: RF8 is used if there is no reasoning revision pair predicted by RC-Evidence. *Path RF-EF3-RF9*: RF9 is used when an unsuccessful revision label is predicted by RC-Evidence and RC-Success. *Path RF-EF3-RF10*: RF10 is provided if not RF8 or RF9. The detailed messages and how they are selected are shown in Table 6 in the Appendix.

## 3.2 eRevise+RF Frontend

eRevise+RF is a web-based system that offers three types of interfaces for students, teachers, and administrators. The student interface (Figure 5 in the Appendix; also see Figure 2) displays the student’s last submitted essay along with its feedback messages. It also allows the student to submit a revised draft. The teacher interface (Figure 6 in the Appendix) includes a submission page that monitors all students’ submitted drafts and feedback messages. It also has a classroom page for retrieving student information. The administrator interface has the highest privilege of accessing teacher and student interfaces as well as managing users (e.g., adding and deleting users).

<sup>3</sup>We use expert-crafted keywords for specific RTA article examples as in Rahimi et al. (2017).



| Student | MVP   |     |                 |           |           |               |             | SPACE |     |                 |           |           |               |            |
|---------|-------|-----|-----------------|-----------|-----------|---------------|-------------|-------|-----|-----------------|-----------|-----------|---------------|------------|
|         | Essay |     | Revision Action |           |           | Revision Type |             | Essay |     | Revision Action |           |           | Revision Type |            |
|         | N     | WC  | Add             | Delete    | Modify    | Surface       | Content     | N     | WC  | Add             | Delete    | Modify    | Surface       | Content    |
| Grade 4 | 165   | 227 | 4.0 / 4.5       | 1.0 / 1.9 | 2.3 / 2.1 | 2.2 / 1.8     | 5.1 / 6.7   | 72    | 176 | 5.3 / 5.2       | 2.4 / 3.2 | 1.9 / 2.1 | 1.7 / 1.8     | 7.8 / 8.7  |
| Grade 5 | 72    | 323 | 9.3 / 5.5       | 6.8 / 2.4 | 3.5 / 3.3 | 2.5 / 2.9     | 17.1 / 8.4  | 129   | 248 | 5.9 / 3.9       | 2.0 / 1.8 | 1.8 / 2.6 | 1.6 / 2.3     | 8.1 / 6.0  |
| Grade 6 | 18    | 324 | 8.0 / 7.8       | 2.8 / 7.7 | 2.3 / 3.0 | 2.8 / 2.2     | 10.3 / 16.3 | -     | -   | -               | -         | -         | -             | -          |
| Grade 7 | 222   | 374 | 4.4 / 2.1       | 1.6 / 0.7 | 2.0 / 1.1 | 1.8 / 1.1     | 6.2 / 2.7   | 225   | 337 | 4.5 / 3.3       | 1.6 / 1.6 | 2.6 / 2.0 | 2.5 / 1.8     | 6.2 / 5.0  |
| Grade 8 | 39    | 399 | 3.5 / 2.5       | 1.5 / 0.9 | 1.1 / 1.8 | 1.5 / 1.4     | 4.5 / 3.8   | 24    | 363 | 3.1 / 6.8       | 1.4 / 8.3 | 2.1 / 2.8 | 2.1 / 2.6     | 4.5 / 15.1 |
| Overall | 516   | 320 | 5.0 / 3.6       | 2.2 / 1.6 | 2.2 / 1.8 | 2.0 / 1.7     | 7.4 / 5.4   | 450   | 287 | 5.0 / 3.9       | 1.8 / 2.3 | 2.2 / 2.2 | 2.1 / 2.0     | 6.9 / 6.4  |

Table 1: The statistics of the collected MVP and SPACE essays in terms of grades, number of essays (N), average word counts (WC) in an essay, and average numbers of revision actions and revision types between draft1-draft2 (before the slash) and draft2-draft3 (after the slash), respectively.

### 3.3 Implementation Detail

The eRevise+RF system is developed based on Django-4.1 and MySQL-8.4. The revision classifiers are implemented with PyTorch<sup>4</sup>, Huggingface<sup>5</sup>, and Azure OpenAI API<sup>6</sup> for ChatGPT-3.5-turbo. We use pretrained DistilRoBERTa-base and BERT-base models from Huggingface as text encoders and multilayer perceptrons as classifier layers. The classifiers are optimized using cross-entropy loss with Adam optimizer on a GeForce RTX 3090 GPU. We set the batch size as 16 and the learning rate as 5e-5 with 5-fold cross-validation, where 80% for training and 20% for parameter tuning using the corpora from prior work. We use NLTK<sup>7</sup> and sentence-splitter<sup>8</sup> for preprocessing.

## 4 System Deployment

The eRevise+RF system was deployed in collaboration with 6 teachers and 406 students in grades 4 to 8 during Spring 2024 for the MVP article and Fall 2024 for the SPACE article. The teachers overlapped for the 2 deployments, while the students were different (e.g., no grade 6 students in Fall). The participants were from one school in Pennsylvania (PA), where 35% of the students are students of color, representing 45 different zip code areas, and speaking more than 10 languages at home; two schools in Louisiana (LA), where 98% and 65% enrollment are minorities, respectively, and student reading proficiency rates range from 46% to 47%.

The system was deployed in three sessions administered over 3 to 5 days (see Figure 2). The two deployments received a total of 194 draft1, 194 draft2, and 172 draft3 for MVP and 176 draft1, 176 draft2, and 150 draft3 for SPACE essays since students might not have consistently participated in all three sessions due to absences. Our analysis is based on the essays from students who participated

|          | Evidence Scoring Indicator (QWK) |      | Revision Classifier (F1) |        |        |
|----------|----------------------------------|------|--------------------------|--------|--------|
|          | NPE                              | SPC  | RC-Con                   | RC-Evi | RC-Suc |
| Values   | 0.67                             | 0.82 | 0.96                     | 0.66   | 0.70   |
| Size (N) | 516                              | 516  | 1,525                    | 1,046  | 1,024  |

Table 2: The evaluation of essay-level evidence scoring indicators and sentence-level revision classifiers based on predicted vs. human annotations on MVP essays.

in all three sessions, yielding 3 drafts of 172 MVP essays (516 total) and of 150 SPACE essays (450 total). Table 1 shows the essay revision statistics across grades, where higher-grade students usually have longer essays, and students are inclined to make adding revisions across all grades.

## 5 Results and Analysis

### 5.1 System NLP Evaluation

We use the 516 MVP essays for manual annotation, to evaluate the accuracy of the scoring indicator algorithms. We employed experts to annotate NPE and SPC scores for every sampled essay. The expert annotators participated in prior studies with Kappas of 0.8 and 0.84 for Evidence and Reasoning (Liu et al., 2023), and ICCs of 0.84 and 0.69 for NPE and SPC (Correnti et al., 2022), respectively. Table 2 shows that Quadratic Weighted Kappa (QWK) between system-predicted and human-annotated NPE and SPC are 0.67 and 0.82, respectively. The agreements are substantial and almost perfect, respectively (Landis and Koch, 1977), which suggests the scoring indicators (NPE and SPC) are effective. Also, we investigate the performance of the EF level predictions on the 172 first drafts since this EF is mainly designed for the first drafts. The QWK between the predicted and annotated evidence feedback levels<sup>9</sup> is 0.66, which suggests the AES(+RF) system used to select evidence use feedback levels for students' first drafts is also reasonably effective.

Again, we employed the trained annotators to annotate 172 pairs of first and second MVP essay

<sup>4</sup><https://pytorch.org/>

<sup>5</sup><https://huggingface.co/>

<sup>6</sup><https://azure.microsoft.com/en-us/>

<sup>7</sup><https://www.nltk.org/>

<sup>8</sup><https://pypi.org/project/sentence-splitter/>

<sup>9</sup>We feed expert-annotated NPE and SPC to level selection algorithm (in Sec. 3.1.3) to obtain annotated EF levels.

drafts, to evaluate the performance of the revision classifiers. The annotators annotated *surface* or *content* for 1,525 revision pairs, and *evidence* or *reasoning* labels for 1,046 out of 1,525 extracted revision pairs, given the other 479 revision pairs were labeled with *surface*, and were removed since RC-Evidence only handles content revisions. Moreover, the annotators annotated the RER scheme (Afrin et al., 2020) on the revision pairs based on the prior work (Liu et al., 2023) and obtained 1,024 annotated *successful* and *unsuccessful* labels, given the other 22 revision pairs were labeled with *claim*, which were removed since RC-Success only handles evidence and reasoning revisions. Table 7 in the Appendix shows an annotated essay example.

Using these annotations to evaluate the three revision classifiers yields RC-Content, RC-Evidence, and RC-Success F1 scores of 0.96, 0.66, and 0.70, respectively (shown in Table 2). Precision/recall figures and a grade breakdown are shown in Table 8 in the Appendix, and suggest that RC-Evidence has relatively lower F1 in grade 8 while RC-Success has relatively lower precision in grades 6 and 7. Confusion matrices are shown in Table 9 in the Appendix, and suggest that RC-Evidence is likely to predict evidence as reasoning and RC-Success makes errors mostly on true unsuccessful revisions.

In sum, both the evidence score indicators and revision classifiers are reasonably effective in using NLP to assess student argumentative writing and revisions (RQ1).

## 5.2 Student Performance

Similar to prior RTA studies (Zhang et al., 2019; Correnti et al., 2022), we evaluated changes in essay quality across drafts using the metrics of Word Count (WC), Evidence Score (ES), NPE and SPC. The NPE and SPC are predicted based on the algorithms described earlier in the evidence score indicators (in Sec. 3.1.1). The ES is predicted with a BERT (Devlin et al., 2019) model trained on a previously collected corpus with 0.89 QWK on MVP essays and 0.94 QWK on SPACE essays<sup>10</sup>. Table 3 shows that essay quality improves over three drafts, where for all metrics, higher is better. Specifically, MVP essays have a 24.6% NPE increase and a 38.5% SPC increase, while SPACE essays have a 27.2% NPE increase and a 29.8% SPC increase from the first to the second drafts, re-

<sup>10</sup>The corpus annotated evidence scores on 317 MVP and 109 SPACE essays using a scale of 1 to 4 (low to high).

| Essay   | MVP (N=516) |      |      |       | SPACE (N=450) |      |      |      |
|---------|-------------|------|------|-------|---------------|------|------|------|
|         | WC          | ES   | NPE  | SPC   | WC            | ES   | NPE  | SPC  |
| Draft 1 | 293         | 2.45 | 2.56 | 8.81  | 228           | 2.09 | 3.02 | 6.07 |
| Draft 2 | 361         | 2.81 | 3.19 | 12.21 | 295           | 2.55 | 3.84 | 7.88 |
| Draft 3 | 405         | 2.91 | 3.34 | 13.98 | 338           | 2.85 | 4.09 | 8.84 |

Table 3: Essay quality improvement based on WC and predicted values of ES, NPE, and SPC metrics.

spectively. ES increases over the three drafts, with the greatest increase (14.69% on MVP and 22.0% on SPACE) occurring from the first to the second drafts. These results reveal observable improvement of the essays from the first to the second draft after evidence use feedback, and from the second to the third drafts after revision feedback.

Additionally, we use the same manually annotated 516 student essays as in Table 2 to evaluate the quality changes between drafts, now using gold rather than predicted values. The annotated NPE improves on average from 2.20, 2.79, to 2.93 in the first, second, and third drafts, respectively, while the predicted NPE improves from 2.56, 3.19, to 3.34 (shown in Table 3). Although the system tends to overscore NPE, the inferences with respect to improvement across drafts are still the same (e.g., 26.8% vs. 24.6% improvement from the first to the second drafts). The annotated SPC improves from 9.00, 12.04, to 13.59 across drafts, while the predicted SPC improves from 8.81, 12.21, to 13.98 (shown in Table 3), which suggests using predictions yields the same conclusions as annotations regarding improvement in essay quality.

## 5.3 Feedback Effectiveness

To better understand whether students are responding to the particular feedback message they receive, we investigate both annotated and predicted NPE and SPC changes between two MVP drafts (old drafts before and new drafts after receiving feedback) in Table 4. For the draft1-draft2 essays, the annotated NPE increases 128% when the old drafts receive EF1 (needing more evidence) feedback, which suggests that students do follow the EF1 to add more evidence, given NPE measures the number of pieces of evidence used. Also, the annotated SPC increases 55.0% when the old drafts receive EF2 (needing more specific details), which also confirms students’ essays have an improvement in specificity after revising essays based on EF2. When the old drafts had EF3 (needing more explanation), annotated NPE mostly remains no change ( $\Delta\text{NPE}=+1.98\%$ ) and SPC has slight improvement. This is because the RF guided students to add explanations, not evidence. These observations confirm

| Evidence Feedback |            | Draft1-Draft2 |              |             |              | Draft2-Draft3 |              |             |              |
|-------------------|------------|---------------|--------------|-------------|--------------|---------------|--------------|-------------|--------------|
|                   |            | NPE           | $\Delta$ NPE | SPC         | $\Delta$ SPC | NPE           | $\Delta$ NPE | SPC         | $\Delta$ SPC |
| Annotated NPE/SPC | EF1 (N=74) | 1.09 / 2.49   | +128%        | 4.41 / 9.18 | +108%        | 2.49 / 2.77   | +11.2%       | 9.18 / 11.7 | +27.8%       |
|                   | EF2 (N=11) | 2.64 / 3.18   | +20.5%       | 6.45 / 10.0 | +55.0%       | 3.18 / 2.82   | -11.3%       | 10.0 / 10.5 | +4.50%       |
|                   | EF3 (N=87) | 3.03 / 3.09   | +1.98%       | 12.9 / 14.6 | +13.0%       | 3.09 / 3.08   | -0.00%       | 14.6 / 15.6 | +6.50%       |
| Predicted NPE/SPC | EF1 (N=74) | 1.28 / 2.74   | +114%        | 4.78 / 10.1 | +112%        | 2.74 / 3.05   | +11.3%       | 10.1 / 12.7 | +24.9%       |
|                   | EF2 (N=11) | 3.18 / 3.45   | +8.49%       | 3.09 / 6.00 | +94.2%       | 3.45 / 3.09   | -10.4%       | 6.00 / 7.73 | +28.8%       |
|                   | EF3 (N=87) | 3.56 / 3.53   | -0.84%       | 13.0 / 14.7 | +13.4%       | 3.53 / 3.61   | +2.27%       | 14.7 / 15.9 | +7.75%       |

Table 4: Annotated and predicted NPE/SPC changes between two MVP essay drafts with respect to received evidence feedback levels on the old drafts. The number before the slash is the NPE and SPC on the old drafts and the latter is on the new drafts.  $\Delta$  means the ratio of the changes from the old drafts to the new drafts. Both evidence feedback (from draft1 to draft2) and revision feedback (from draft2 to draft3) help students improve their essays.

that how student revisions change their evidence usage is as expected for all three EF levels. Similarly, we compare draft2-draft3 essays before and after receiving revision feedback. We observe annotated NPE improvement when draft2 has EF1 (e.g., RF was either RF3, RF4, RF5, or RF6 in Figure 3), which suggests revision feedback helps students add more evidence. However, revision feedback may not be helpful when draft2 receives EF2 given the annotated NPE average drops by 11.3%, which suggests that student essays might slightly weaken the completeness of the evidence while revising toward specificity. These observations are consistent when using predicted NPE and SPC in Table 4. Also, we analyze predicted NPE and SPC changes for SPACE essays in Table 10 in the Appendix, which shows similar patterns. In sum, the average increases of NPE and SPC for two RTA essays suggest that both evidence use feedback and revision feedback are helpful for students to improve their writing, in ways that are in alignment with the system’s feedback messages (RQ2).

## 5.4 Case Study

We show a student’s three argumentative essay drafts in Figure 4 to illustrate how the feedback helps the student improve the essay through revision. The first draft does not include sufficient evidence to support the claim as it has predicted NPE of 2. Thus, the student receives EF that “*adding more evidence would make your argument even more convincing.*” Afterward, the student makes surface (typo) and content revisions in the second draft by deleting one piece of evidence about *school* (purple) and adding another piece of evidence about *farming* (yellow). The second draft has the same predicted NPE of 2 but SPC increases from 1 to 5 since it replaces *school* with *farming* and adds more details about *farming*. Additional RF is provided “*...When writers revise their text-based essays, they generally add new content from the text and delete content that is not based on the text.*” In response,

|  |
|--|
| <b>Draft1 (NPE=2, SPC=1):</b> ... The people only get payed \$1 every day wich is why 20,000 people die from malaria every day! they can't pay for a cure. Also, because of the money, the schools can't pay for meals and they only have a limited amount of school supplies for the students to use. I think all of this is possible to fix,if every one worked harder to solve these problems.  |
| <b>Draft2 (NPE=2, SPC=5):</b> ... The people only get payed \$1 every day <del>wich</del> <b>which</b> is why 20,000 people die from malaria every day! they can't pay for a cure. <del>Also, because of the money, the schools can't pay for meals and they only have a limited amount of school supplies for the students to use.</del> Another problem is the farmers can't pay for fertilizer or water for their plants. "each year, the farmers worry: will they harvest enough food to feed the whole family? Will their kids go hungry and become sick?" I think all of this is possible to fix,if every one worked harder to solve these problems.   |
| <b>Draft3 (NPE=3, SPC=9):</b> ... The people only get payed \$1 every day which is why 20,000 people die from malaria every day! they can't pay for a cure. Another problem is the farmers can't pay for fertilizer or water for their plants. "each year, the farmers worry: will they harvest enough food to feed the whole family? Will their kids go hungry and become sick?" <b>Another problem is at the hospital there are only three kids to a bed and two adults to a bed. "the rooms were packed with patients who probably should not receive treatment, either because the hospital did not have it or the patients could not afford it."</b> I think all of this is possible to fix, if we all worked a <del>little bit</del> <b>little bit</b> harder to fix these problems. |

Figure 4: The example of a MVP essay, revising from draft1 to draft2, then from draft2 to draft3. The purple marks deletion, yellow and green marks addition.

the student makes additional surface and content revisions in the third draft by adding a piece of evidence about *hospital* with more details, which increases predicted NPE to 3 and SPC to 9. This example suggests the feedback is helpful for the student to reflect on revisions that have been made and further revise the essay. We show additional cases in Table 11 and 12 in the Appendix.

## 6 Conclusion

We present eRevise+RF, an advanced version of the eRevise system that extends and integrates algorithms to assess argumentative essay revisions, and to provide formative feedback on both evidence usage and revision success in response to feedback. Although the algorithms are primarily adapted from our prior published work, their integration into a complete and deployed system is a novel contribution. Through our pilot studies, we show promise of using NLP algorithms to scaffold young students in writing and revising text-based argumentative essays. In future work, we will deploy the system in more classrooms and with other RTA articles, to generalize our findings.

## Ethics Statement

eRevise+RF was used to collect our essay corpus under standard protocols approved by an institutional review board (IRB). Our data is not publicly available at the moment to ensure the safety of the private information of young students. The system is currently accessible only to the teachers and students participating in our study who have been assigned user accounts, in compliance with IRB constraints and data privacy requirements that restrict open user registration. However, the system source code is made publicly available on GitHub to support contributions to the NLP community. The system does not pose any ethical concerns because of the limited access to the data, but there might be a risk that the system may give poor advice based on incorrect scoring indicators or classifier predictions, or that the predictive models may learn biases due to small annotated training corpora. However, due to restrictions of IRB and policies of partner schools, only school-level but not student-level demographics are available which makes it difficult to evaluate potential biases.

## Limitations

**System.** Although the eRevise+RF system demonstrates promise in helping students revise essays in response to formative feedback, its evaluation is based on a relatively small essay corpus (i.e., 516 MVP and 450 SPACE essays). Nonetheless, the corpus exhibits substantial diversity in both grade levels and demographics. For example, one school in PA includes students from 45 different zip code areas, while two schools in LA have minority enrollments of 98% and 65%, respectively. The analyses in Sec. 5.2 demonstrate the robustness of students’ essay improvement across these diverse settings. Although the system deployments were pilot and lacked a control group where students wrote without receiving feedback since we didn’t have enough classrooms to make this feasible, it is a goal of future studies. However, our analysis in Sec. 5.3 was designed to start teasing apart the impact of the system on students compared to unscaffolded revision. Our results show that the student revisions were in fact largely responsive to the feedback. Although we currently only assess the evidence dimension of the RTA, we started there because a persistent criticism of AWE systems is that they have not been designed to meet ambitious writing standards such as text-

based argument writing (Burstein et al., 2020). But, we note other aspects of writing quality such as creativity need to be studied in future work.

**LLM.** Our methods are mostly based on rubrics designed by RTA experts rather than advanced LLM. This is because we notice that for our deployment with young students, hallucinations and other LLM generation concerns such as offensive content and biases would need to be carefully studied and addressed. Additionally, a recent study (Behzad et al., 2024) compares LLM-generated essay feedback with human feedback and concludes that human feedback is more effective in delivering specific and actionable comments that address the most critical issues in an essay. This suggests young student learners may benefit more from expert feedback tailored to their specific writing challenges. In future work, we aim to integrate both expert and LLM-generated feedback into the system to provide nuanced and adaptive feedback messages.

**Algorithm.** The evidence scoring indicators described in Sec 3.1.1 sometimes overcount because the text-based evidence is not perfectly captured. Since the initial deployment, the collected and annotated essays have been used to tune the sliding window and text similarity algorithms in a data-driven manner. We have since modified the NPE algorithm based on an error analysis and it has now achieved a QWK of 0.87 compared to the original 0.67 in Table 2. The scoring indicator’s reliance on manual keywords may not adapt well to different writing topics or styles unless continuously updated. Thus, other methods (Zhang and Litman, 2020, 2021) to avoid dependence on manual keyword identification could be implemented. Also, the trained revision classifiers RC-Evidence and RC-Success are not excellent, thus may cause the AES+RF system to provide incorrect revision feedback, which could be improved in future work.

## Acknowledgments

This research was supported by National Science Foundation Award #2202347 and its REU and NAIRR Pilot Demonstration Project supplements, as well as through CloudBank allocations. The opinions expressed are those of the authors and do not represent the views of the institutes. The authors would like to thank the anonymous reviewers, and the Pitt PETAL and NLP groups for their valuable feedback on this work.



## References

- Tazin Afrin, Elaine Lin Wang, Diane Litman, Lindsay Clare Matsumura, and Richard Correnti. 2020. Annotation and classification of evidence and reasoning revisions in argumentative writing. In *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*, Seattle, Washington, USA (Remote).
- Talita Anthonio, Irshad Bhat, and Michael Roth. 2020. [wikiHowToImprove: A resource and analyses on edits in instructional texts](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5721–5729, Marseille, France. European Language Resources Association.
- Shabnam Behzad, Omid Kashefi, and Swapna Soma-sundaran. 2024. [Assessing online writing feedback resources: Generative AI vs. good samaritans](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 1638–1644, Torino, Italia. ELRA and ICCL.
- Jill Burstein, Brian Riordan, and Daniel McCaffrey. 2020. Expanding automated writing evaluation. In *Handbook of automated scoring*, pages 329–346. Chapman and Hall/CRC.
- Ruining Chong, Cunliang Kong, Liu Wu, Zhenghao Liu, Ziyi Jin, Liner Yang, Yange Fan, Hanghang Fan, and Erhong Yang. 2023. [Leveraging prefix transfer for multi-intent text revision](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1219–1228, Toronto, Canada. Association for Computational Linguistics.
- Richard Correnti, Lindsay Clare Matsumura, Laura Hamilton, and Elaine Wang. 2013. Assessing students’ skills at writing analytically in response to texts. *The Elementary School Journal*, 114(2):142–177.
- Richard Correnti, Lindsay Clare Matsumura, Elaine Wang, Diane Litman, Zahra Rahimi, and Zahid Kisa. 2020. Automated scoring of students’ use of text evidence in writing. *Reading Research Quarterly*, 55(3):493–520.
- Richard Correnti, Lindsay Clare Matsumura, Elaine Lin Wang, Diane Litman, and Haoran Zhang. 2022. [Building a validity argument for an automated writing evaluation system \(revise\) as a formative assessment](#). *Computers and Education Open*, 3:100084.
- Rip Correnti, Elaine Lin Wang, Lindsay Clare Matsumura, Diane Litman, Zhexiong Liu, and Tianwen Li. 2024. Supporting students’ text-based evidence use via formative automated writing and revision assessment. In *The Routledge international handbook of automated essay evaluation*, pages 221–243. Routledge.
- Mike D’Arcy, Alexis Ross, Erin Bransom, Bailey Kuehl, Jonathan Bragg, Tom Hope, and Doug Downey. 2024. [ARIES: A corpus of scientific paper edits made in response to peer reviews](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6985–7001, Bangkok, Thailand. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Wanyu Du, Vipul Raheja, Dhruv Kumar, Zae Myung Kim, Melissa Lopez, and Dongyeop Kang. 2022. [Understanding iterative revision from human-written text](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3573–3590, Dublin, Ireland. Association for Computational Linguistics.
- Johanna Fleckenstein, Lucas W Liebenow, and Jennifer Meyer. 2023. Automated feedback and writing: A multi-level meta-analysis of effects on students’ performance. *Frontiers in Artificial Intelligence*, 6:1162454.
- Steve Graham, Karen R Harris, and Tanya Santangelo. 2015. Research-based writing practices and the common core: Meta-analysis and meta-synthesis. *The Elementary School Journal*, 115(4):498–522.
- Qian Guo, Ruiling Feng, and Yuanfang Hua. 2023. *Automated written corrective feedback in Research Paper Revision: The good, the bad, and the missing*. Routledge.
- Shi Huawei and Vahid Aryadoust. 2023. A systematic review of automated writing evaluation systems. *Education and Information Technologies*, 28(1):771–795.
- Léane Jourdan, Florian Boudin, Nicolas Hernandez, and Richard Dufour. 2024. [CASIMIR: A corpus of scientific articles enhanced with multiple author-integrated revisions](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 2883–2892, Torino, Italia. ELRA and ICCL.
- Omid Kashefi, Tazin Afrin, Meghan Dale, Christopher Olshefski, Amanda Godley, Diane Litman, and Rebecca Hwa. 2022. [Argrewrite v.2: an annotated argumentative revisions corpus](#). *Language Resources and Evaluation*, pages 1574–0218.
- Deanna Kuhn, Laura Hemberger, and Valerie Khait. 2017. *Argue with me: Argument as a path to developing students’ thinking and writing*. Routledge.

- J. Richard Landis and Gary Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- Tianwen Li, Zhexiong Liu, Lindsay Matsumura, Elaine Wang, Diane Litman, and Richard Correnti. 2024. [Using large language models to assess young students’ writing revisions](#). In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 365–380, Mexico City, Mexico. Association for Computational Linguistics.
- Diane Litman, Tazin Afrin, Omid Kashefi, Christopher Olshefski, Amanda Godley, and Rebecca Hwa. 2022. [An automated writing evaluation system for supporting self-monitored revising](#). In *Artificial Intelligence in Education: 23rd International Conference, AIED 2022, Durham, UK, July 27–31, 2022, Proceedings, Part I*, page 581–587, Berlin, Heidelberg. Springer-Verlag.
- Lei Liu and Min Zhu. 2022. Bertalign: Improved word embedding-based sentence alignment for chinese–english parallel corpora of literary texts. *Digital Scholarship in the Humanities*.
- Zhexiong Liu, Diane Litman, Elaine Wang, Lindsay Matsumura, and Richard Correnti. 2023. [Predicting the quality of revisions in argumentative writing](#). In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 275–287, Toronto, Canada. Association for Computational Linguistics.
- Masato Mita, Keisuke Sakaguchi, Masato Hagiwara, Tomoya Mizumoto, Jun Suzuki, and Kentaro Inui. 2024. [Towards automated document revision: Grammatical error correction, fluency edits, and beyond](#). In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 251–265, Mexico City, Mexico. Association for Computational Linguistics.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. [GloVe: Global vectors for word representation](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.
- Zahra Rahimi, Diane Litman, Richard Correnti, Elaine Wang, and Lindsay Clare Matsumura. 2017. Assessing students’ use of evidence and organization in response-to-text writing: Using natural language processing for rubric-based automated scoring. *International Journal of Artificial Intelligence in Education*, 27(4):694–728.
- Rod D. Roscoe, Erica L. Snow, and Danielle S. McNamara. 2013. Feedback and revising in an intelligent tutoring system for writing strategies. In *Artificial Intelligence in Education*, pages 259–268, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. [Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter](#). In *EMC2: Efficient Methods for Deep Learning Workshop at NeurIPS 2019*.
- Antonette Shibani, Simon Knight, and Simon Buckingham Shum. 2018. Understanding revisions in student writing through revision graphs. In *International Conference on Artificial Intelligence in Education*, pages 332–336, Cham. Springer International Publishing.
- Alexander Spangher, Xiang Ren, Jonathan May, and Nanyun Peng. 2022. [NewsEdits: A news article revision dataset and a novel document-level reasoning challenge](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 127–157, Seattle, United States. Association for Computational Linguistics.
- Elaine Lin Wang, Lindsay Clare Matsumura, Richard Correnti, Diane Litman, Haoran Zhang, Emily Howe, Ahmed Magooda, and Rafael Quintana. 2020. [erevis\(ing\): Students’ revision of text evidence use in an automated writing evaluation system](#). *Assessing Writing*, 44:100449.
- Joshua Wilson, Yue Huang, Corey Palermo, Gaysha Beard, and Charles A MacArthur. 2021. Automated feedback and automated scoring in the elementary grades: Usage, attitudes, and associations with writing outcomes in a districtwide implementation of mi write. *International Journal of Artificial Intelligence in Education*, 31(2):234–276.
- Fan Zhang, Homa B. Hashemi, Rebecca Hwa, and Diane Litman. 2017. [A corpus of annotated revisions for studying argumentative writing](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1568–1578, Vancouver, Canada. Association for Computational Linguistics.
- Fan Zhang, Rebecca Hwa, Diane Litman, and Homa B Hashemi. 2016. Argrewrite: A web-based revision assistant for argumentative writings. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Demonstrations*, pages 37–41.
- Haoran Zhang and Diane Litman. 2020. [Automated topical component extraction using neural network attention scores from source-based essay scoring](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8569–8584, Online. Association for Computational Linguistics.
- Haoran Zhang and Diane Litman. 2021. [Essay quality signals as weak supervision for source-based essay scoring](#). In *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 85–96, Online. Association for Computational Linguistics.

Haoran Zhang, Ahmed Magooda, Diane Litman, Richard Correnti, Elaine Wang, LC Matsmura, Emily Howe, and Rafael Quintana. 2019. *erevise: Using natural language processing to provide formative feedback on text evidence usage in student writing*. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9619–9625.

Timon Ziegenbein, Gabriella Skitalinskaya, Alireza Bayat Makou, and Henning Wachsmuth. 2024. [LLM-based rewriting of inappropriate argumentation using reinforcement learning from machine feedback](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4455–4476, Bangkok, Thailand. Association for Computational Linguistics.

## A MVP Article

### *A Brighter Future*

Hannah Sachs

*The unpaved dirt road made our car jump as we traveled to the Millennium Village in Sauri (sah-ooh-ree), Kenya. We passed the market where women sat on the dusty ground selling bananas. Little kids were wrapped in cloth on their mothers' backs, or running around in bare feet and tattered clothing. When we reached the village, we walked to the Bar Sauri Primary School to meet the people. Welcoming music and singing had almost everyone dancing. We joined the dancing and clapped along to the joyful, lively music.*

*The year was 2014, the first time I had ever been to Sauri. With the help of the Millennium Villages project, the place would change dramatically in the coming years. The Millennium Villages project was created to help reach the Millennium Development Goals. The plan is to get people out of poverty, assure them access to health care and help them stabilize the economy and quality of life in their communities. Villages get technical advice and practical items, such as fertilizer, medicine and school supplies. Local leaders take it from there. The goals are supposed to be met by 2025; some other targets are set for 2035. We are halfway to 2035, and the world is capable of meeting these goals. But our first glimpse of Sauri showed us that there was plenty of work to do.*

### *The Fight for Better Health*

*On that day in 2014, we followed the village leaders into Yala Sub-District Hospital. It was not in good shape. There were three kids to a bed and two adults to a bed. The rooms were packed with*

*patients who probably would not receive treatment, either because the hospital did not have it or the patients could not afford it. There was no doctor, only a clinical officer running the hospital. There was no running water or electricity. It is hard for me to see people sick with preventable diseases people who are near death when they shouldn't have to be. I just get scared and sad.*

*Malaria (mah-lair-eeh-ah) is one disease, common in Africa, that is preventable and treatable. Mosquitoes carry malaria, and infect people by biting them. Kids can die from it easily, and adults get very sick. Mosquitoes that carry malaria come at night. A bed net, treated with chemicals that last for five years, keeps malarial mosquitoes away from sleeping people. Each net costs \$5. There are some cheap medicines to get rid of malaria too. The solutions are simple, yet 20,000 kids die from the disease each day. So sad, and so illogical. Bed nets could save millions of lives.*

### *Water, Fertilizer, Knowledge*

*We walked over to see the farmers. Their crops were dying because they could not afford the necessary fertilizer and irrigation. Time and again, a family will plant seeds only to have an outcome of poor crops because of lack of fertilizer and water. Each year, the farmers worry: Will they harvest enough food to feed the whole family? Will their kids go hungry and become sick?*

*Many kids in Sauri did not attend school because their parents could not afford school fees. Some kids are needed to help with chores, such as fetching water and wood. In 2014, the schools had minimal supplies like books, paper and pencils, but the students wanted to learn. All of them worked hard with the few supplies they had. It was hard for them to concentrate, though, as there was no midday meal. By the end of the day, kids didn't have any energy.*

### *A Better Life-2018*

*The people of Sauri have made amazing progress in just four years. The Yala Sub-District Hospital has medicine, free of charge, for all of the most common diseases. Water is connected to the hospital, which also has a generator for electricity. Bed nets are used in every sleeping site in Sauri. The hunger crisis has been addressed with fertilizer and seeds, as well as the tools needed to maintain the food supply. There are no school fees, and the school now serves lunch for the students. The attendance rate is way up.*

*Dramatic changes have occurred in 80 villages across sub-Saharan Africa. The progress is encouraging to supporters of the Millennium Villages project. There are many solutions to the problems that keep people impoverished. What it will really take is for the world to work together to change poverty-stricken areas for good. When my kids are my age, I want this kind of poverty to be a thing of history. It will not be an easy task. But Sauri's progress shows us all that winning the fight against poverty is achievable in our lifetime.*

#### Essay Prompt

*The author described how the quality of life can be improved by the Millennium Villages project in Sauri, Kenya. Based on the article, did the author convince you that "winning the fight against poverty is achievable in our lifetime"? Explain why or why not with 3 to 4 examples from the text to support your answer.*

## B SPACE Article

### *The Importance of Space Exploration*

*Is space exploration really worthwhile when so much needs to be done on Earth? This is a question that has been asked for decades and requires serious thought.*

### *Arguments Against Space Exploration*

*The arguments against space exploration stem from a belief that the money spent could be used differently – to improve people's lives. In 1953, President Eisenhower captured this viewpoint. He opposed the space program, saying that each rocket fired was a theft from citizens who suffer from hunger and poverty. Indeed, over 46.2 million Americans (15%) live in Poverty. Nearly half of all Americans also have trouble paying for basic needs such as housing, food, or medicine at some point in their lives.*

*In other countries people are dying because they do not have access to clean water, medical care or simple solutions that prevent the spread of diseases. For example, malaria, a disease spread by mosquito bites kills over 3,000 African children every day. This could be prevented simply by hanging large nets over children's beds. This would protect people from being bitten as they sleep. These nets cost only \$5 each, but most families at risk for malaria cannot afford them.*

*It is not just people that need help. The Earth is suffering also. Many scientists believe that pol-*

*lution from burning fossil fuels (gasoline and oil) is destroying our air and oceans. We need new, cleaner forms of energy to power cars, homes, and factories. A program to develop clean energy could be viewed as a worthy investment. Maybe exploring space should not be a priority when there is so much that needs to be done on Earth. In 2012, the United States spent 19 billion dollars for space exploration. Some people think that money should be spent to help heal the people and the Earth.*

### *Benefits of Space Exploration*

*Those in favor of space exploration argue that 19 billion dollars is not too much. Nineteen billion was only 1.2% of the total national budget. This is a tiny amount compared to the 670 billion spent that year on national defense (26.3% of the national budget). It is even less than the 70 billion the federal government spent on education (48% of the budget).*

*Another reason to spend money on space exploration is that it has led to benefits in many fields. One such field is medicine. Before NASA could send astronauts into space, scientists needed to find ways to monitor their health under stressful conditions. NASA wanted to make sure the astronauts could survive the harsh conditions of launch and reentry. The scientists developed medical instruments to monitor body functions. They also learned a lot about how the human body reacts to stress. In rising to meet the challenges of space exploration, NASA scientists have developed other innovations that improve our lives. These include better exercise machines and airplanes, and more accurate weather forecasting. All of these innovations resulted from technologies that engineers developed to make space travel possible.*

*Even the problems of hunger and poverty can be tackled by space exploration. Satellites that circle Earth can monitor land and the atmosphere. They can track and measure the conditions of crops, soil, and rainfall. We can use this information to improve the way we produce and distribute food. This will enable us to provide more food at a lower price to people who need it. When we explore space, we are also helping to solve serious problems on Earth.*

*Beyond providing us with inventions, space exploration is important for the challenge it provides and the motivation to bring out the best in ourselves. Often we make progress in solving difficult problems by first setting challenging goals, which*



*inspire innovative work. Space exploration is important because it can motivate beneficial competition among nations. Imagine how much human suffering can be avoided if nations competed with planet-exploring spaceships instead of bomb-dropping airplanes. We saw an example of this during the 1960s. During the Cold War, the United States and Russia competed in a race to explore space. They each wanted to be the first to land a spacecraft on the moon and visit other planets. The National Academy of Science says that this competition led to significant investments and improvements in American education, especially in math and science. This shows that by looking outward into space, we also improve life here on Earth.*

*All this brings us back to the question: Should we explore space when there is so much that needs to be done on Earth? It is true that we have many serious problems to deal with on Earth. However, space exploration is not at odds with solving human problems. In fact, it may even help find solutions. Space exploration leads to long-term benefits that more than justify the immediate cost.*

#### *Essay Prompt*

*Consider the reasons given in the article for why we should and should not fund space exploration. Did the author convince you that “space exploration leads to long-term benefits” that justify the cost? Give reasons for your answer. Support your reasons with 3 to 4 pieces of evidence from the text.*

| Level | Assessment   | Name                        | Feedback   |
|-------|--|-----------------------------|--|
| EF1   | NPE less or equal to $\alpha$                            | Needs more evidence         | <ul style="list-style-type: none"> <li>•Adding more evidence would make your argument even more convincing.</li> <li>•Reread the highlighted portions of the article to choose more evidence (only use if texts were highlighted).</li> </ul>  |
| EF2   | NPE greater than $\alpha$ , SPC less or equal to $\beta$ | Needs more specific details | <ul style="list-style-type: none"> <li>•Adding more details will help your reader better understand your ideas. This will make your argument even more convincing.</li> <li>•When you revise your essay, make sure to add more details for each piece of evidence you use.</li> </ul>  |
| EF3   | NPE greater than $\alpha$ , SPC greater than $\beta$     | Needs more explanation      | <ul style="list-style-type: none"> <li>•Having evidence is important, but you need to help your reader understand how the evidence you chose supports your argument.</li> <li>•When you revise your essay, focus on explaining how each piece of evidence you used connects to your idea.</li> <li>•Give a detailed and clear explanation of how the evidence supports your argument.</li> <li>•Tie the evidence not only to the point you are making within a paragraph, but to your overall argument.</li> </ul> |

Table 5: Selecting evidence use (EF) feedback levels using NLP-based assessment.

| Level | Assessment   | Name   | Feedback  |
|-------|--|--|---|
| RF1   | No content and no surface revisions or all deletions                           | No attempt                                     | <ul style="list-style-type: none"> <li>•When writers revise, they generally add more content. This often makes their essays longer.</li> <li>•This time when you revise your essay, focus on adding more evidence.</li> </ul>   |
| RF2   | No content revision  | surface revision                               | <ul style="list-style-type: none"> <li>•When you revised your essay, it looks like you edited your writing to be clearer and easier for a reader to understand.</li> <li>•Revising is different from editing. When writers revise their essays, they generally add more content. This often makes their essays longer.</li> <li>•This time when you revise your essay, focus on adding more evidence.</li> </ul>  |
| RF3   | No NPE change; Added sentences contain topic words already used                | Repeats evidence                               | <ul style="list-style-type: none"> <li>•When you revised your essay, it looks like you added in evidence that was very similar to the evidence you had included before.</li> <li>•When writers revise, they generally add new content to their essays.</li> </ul>   |
| RF4   | No NPE change; Added sentences contain no topic words                          | Added evidence but not text based              | <ul style="list-style-type: none"> <li>•When you revised your essay, it looks like you added more information about your thinking but did not include new information from the article.</li> <li>•When writers revise their text-based essays, they generally add new content from the text and delete content that is not based on the text.</li> </ul>  |
| RF5   | NPE change (new NPE >old NPE) and SPC changes less or equal to $\gamma$        | Added evidence but vague or general            | <ul style="list-style-type: none"> <li>•When you revised your essay, it looks like you followed the suggestion to add more evidence. Great job!</li> <li>•When writers revise, they don't just add more information. They also add more details to the information they already have in their essay. This often makes their essays longer.</li> </ul>   |
| RF6   | SPC changes less or equal to $\gamma$ and added sentences contain no SPC words | Added evidence successfully                    | <ul style="list-style-type: none"> <li>•When you revised your essay, it looks like you followed the suggestion to add more evidence. Great job!</li> <li>•Paying attention to feedback is how people become stronger writers.</li> </ul>  |
| RF7   | SPC changes greater than $\gamma$ (new SPC >old SPC)                           | Added specific details successfully            | <ul style="list-style-type: none"> <li>•When you revised your essay, it looks like you followed the suggestion to add more details to your essay. Great job!</li> <li>•Paying attention to feedback is how people become stronger writers.</li> </ul>   |
| RF8   | No reasoning revision  | Successful evidence but no reasoning attempt   | <ul style="list-style-type: none"> <li>•When you revised your essay, it looks like you may have focused on something other than explaining your evidence.</li> <li>•Revising the explanation or reasoning part of an essay is hard to do! When writers revise for this, they make sure that after providing a piece of evidence, they say something that connects it to their argument. The explanation should not just restate the evidence in different words.</li> </ul> |
| RF9   | Unsuccessful reasoning revision  | Successful evidence but unsuccessful reasoning | <ul style="list-style-type: none"> <li>•When you revised your essay, it looks like you may have focused on something other than explaining your evidence.</li> <li>•Revising the explanation or reasoning part of an essay is hard to do! When writers revise for this, they make sure that after providing a piece of evidence, they say something that connects it to their argument. The explanation should not just restate the evidence in different words.</li> </ul> |
| RF10  | Successful reasoning revision  | successful evidence and successful reasoning   | <ul style="list-style-type: none"> <li>•When you revised your essay, it looks like you followed the suggestion to explain your evidence and how it connects to your claim. Great job!</li> <li>•Paying attention to feedback is how people become stronger writers.</li> </ul>  |

Table 6: Selecting revision feedback levels using NLP-based assessments.

| ID | Original Draft Sentence  | Revised Draft Sentence  | Revision Action | Revision Type | Revision Purpose        | Quality Label |
|----|--|---|-----------------|---------------|-------------------------|---------------|
| 0  | In the article "a brighter future", the author proved to me that it is possible to end poverty in a lifetime.        | In the article "a brighter future", the author proved to me that it is possible to end poverty in a lifetime.   | N/A             | N/A           | N/A                     | N/A           |
| 1  | She did this by showing how other countries helped them, how they got more food, and how their water supply went up. | She did this by showing how when they worked together they made medication more affordable (if not free), got more food, and got their water supply up. | Modify          | Content       | Relevant Evidence       | Successful    |
| 2  | ...  | ...   | ...             | ...           | ...                     | ...           |
| 3  | Although some countries are having wars right now, we are still one big team.  |   | Delete          | Content       | not Text-based Evidence | Successful    |
| 4  |  | First, they did it by making medication more affordable.  | Add             | Content       | Relevant Evidence       | Successful    |
| 5  | ...  | ...   | ...             | ...           | ...                     | ...           |
| 6  |  | It is most common in Africa.  | Add             | Content       | not Text-based Evidence | Unsuccessful  |
| 7  | I know this because the author said, "water is connected to the hospital.  | I know this because the author said, "water is connected to the hospital."  | Modify          | Surface       | N/A                     | N/A           |
| 8  | Water is critical to life, so if there is a shortage then no wonder people were dying.                               | Water is critical to life, so if there is a shortage then no wonder people were dying.  | N/A             | N/A           | N/A                     | N/A           |
| 9  | ...  | All of these reasons make it clear that if you put your mind to something, you can achieve it.  | Add             | Content       | LCE Reasoning           | Successful    |
| 10 | ...  | ...   | ...             | ...           | ...                     | ...           |
| 11 | Life is hard, but dying is harder.   | Life is hard, but you don't have to work alone on hard things.  | Modify          | Content       | Commentary Reasoning    | Unsuccessful  |

Table 7: Example of revision annotation between two essay drafts. The sentence alignment (revision action) and classification (revision type) were first done by the system and then manually justified and corrected by annotators. The revision purpose and quality label are annotated based on the RER scheme (Afrin et al., 2020; Liu et al., 2023).

|         | RC-Content |        |          | RC-Evidence |        |          | RC-Success |        |          |
|---------|------------|--------|----------|-------------|--------|----------|------------|--------|----------|
|         | Precision  | Recall | F1-Score | Precision   | Recall | F1-Score | Precision  | Recall | F1-Score |
| Grade 4 | 0.90       | 0.92   | 0.91     | 0.85        | 0.53   | 0.65     | 0.64       | 0.87   | 0.74     |
| Grade 5 | 0.99       | 0.97   | 0.98     | 0.78        | 0.59   | 0.67     | 0.78       | 0.84   | 0.81     |
| Grade 6 | 0.93       | 0.97   | 0.95     | 0.89        | 0.55   | 0.68     | 0.47       | 1.00   | 0.64     |
| Grade 7 | 0.97       | 0.99   | 0.98     | 0.77        | 0.59   | 0.67     | 0.43       | 0.93   | 0.59     |
| Grade 8 | 0.98       | 0.98   | 0.98     | 0.56        | 0.56   | 0.56     | 0.65       | 0.90   | 0.76     |
| Overall | 0.96       | 0.97   | 0.96     | 0.78        | 0.57   | 0.66     | 0.58       | 0.88   | 0.70     |

Table 8: The grade-level evaluation of revision classifiers based on predicted values vs. human annotations. The RC-Content classifier has good performance on all grades, while RC-Evidence has relatively lower results in grade 8 and the precision in grades 6 and 7 are relatively low in RC-Success.

| RC-Content |         |         | RC-Evidence |           |          | RC-Success |           |         |
|------------|---------|---------|-------------|-----------|----------|------------|-----------|---------|
|            | Surface | Content |             | Reasoning | Evidence |            | Unsuccess | Success |
| Surface    | 268     | 50      | Reasoning   | 318       | 99       | Unsuccess  | 100       | 358     |
| Content    | 37      | 1170    | Evidence    | 269       | 360      | Success    | 67        | 499     |

Table 9: The confusion matrices for revision classifiers based on human annotations (rows) vs. predicted values (columns). The RC-Evidence is likely to predict evidence as reasoning and RC-Success makes errors mostly on true unsuccessful revisions.

| Evidence Feedback | Draft1-Draft2 |              |           |              | Draft2-Draft3 |              |           |              |
|-------------------|---------------|--------------|-----------|--------------|---------------|--------------|-----------|--------------|
|                   | NPE           | $\Delta$ NPE | SPC       | $\Delta$ SPC | NPE           | $\Delta$ NPE | SPC       | $\Delta$ SPC |
| EF1 (N=55)        | 1.05/2.71     | +158%        | 1.6/4.91  | +207%        | 2.71/3.27     | +20.7%       | 4.91/6.60 | +34.4%       |
| EF2 (N=49)        | 3.35/3.84     | +14.6%       | 5.20/6.67 | +28.3%       | 3.84/4.02     | +4.69%       | 6.67/7.55 | +13.2%       |
| EF3 (N=46)        | 5.02/5.20     | +3.59%       | 12.3/12.7 | +3.16%       | 5.20/5.15     | -0.96%       | 12.7/12.9 | +1.34%       |

Table 10: Predicted NPE and SPC changes between two SPACE essay drafts with respect to received evidence feedback levels on the old drafts. The number before the slash is the NPE and SPC on the old drafts and the latter is on the new drafts.  $\Delta$  means the ratio of the changes from the old drafts to the new drafts. Both evidence feedback (from draft1 to draft2) and revision feedback (from draft2 to draft3) help students improve their essays.

| Draft1   | Draft2  | Draft3  |
|--|---|---|
| <b>Evidence Scoring Indicators:</b> NPE=1, SPC=5   | <b>Evidence Scoring Indicators:</b> NPE=3, SPC=10   | <b>Evidence Scoring Indicators:</b> NPE=3, SPC=21   |
| <b>Evidence Feedback:</b> Adding more evidence would make your argument even more convincing. Reread the highlighted portions of the article to choose more evidence.  | <b>Revision Feedback:</b> When you revised your essay, it looks like you added more information about your thinking but did not include new information from the article. When writers revise their text-based essays, they generally add new content from the text and delete content that is not based on the text.   | N/A   |
| I believe that we can win the fight against poverty because we are all helping each other to make a difference each day. The text states that the targets are set for 2035. This means that we are moving quickly and we hope to be done soon. The progress over four years was incredibly fast. The hospital had medicine and there are no school fees. This shows the progress is climbing higher on the scale each day. "Dramatic changes have occurred in 80 villages across sub-Saharan Africa". Therefore, the Millennium Villages have gotten more than I could simply say that poverty will soon be gone for good. The text is about how the Millennium Villages program is taking such a huge lead against poverty. They have a good chance of winning and I think they will. | I believe that we can win the fight against poverty because we are all helping each other to make a difference each day. The text states that the targets are set for 2035. This means that we are moving quickly and we hope to be done soon. The progress over four years was incredibly fast. The hospital had medicine and there are no school fees. This shows the progress is climbing higher on the scale each day. "Dramatic changes have occurred in 80 villages across sub-Saharan Africa". Therefore, the Millennium Villages have gotten more than I could simply say that poverty will soon be gone for good. The disease, Malaria, was killing 20,000 kids a day because the parents could not afford bed nets. Yet now, "bed nets are used in every sleeping site in Sauri". The text is about how the Millennium Villages program is taking such a huge lead against poverty. They have a good chance of winning and I think believe that if we all help them, they will win.   | I believe that we can win the fight against poverty because we are all helping each other to make a difference each day. The text states that the targets are set for 2035. This means that we are moving quickly and we hope to be done soon. The progress over four years was incredibly fast. The hospital had medicine and there are no school fees. This shows the progress is climbing higher on the scale each day. This means that we are moving quickly and we hope to be done soon. In the text, it says "Dramatic changes have occurred in 80 villages across sub-Saharan Africa". Therefore, the Millennium Villages have gotten more than done so much that I could simply say that poverty will soon be gone for good. The disease, Malaria, was killing 20,000 kids a day because the parents could not afford bed nets. Yet now, "bed nets are used in every sleeping site in Sauri". All the farmers worry: "Will they harvest enough food to feed the whole family? Will their kids go hungry and become sick?" Now the farmers have the proper needs to grow strong healthy plants. Many kids in Sauri did not go to school. Their parents could not afford the school fees. Now there are no school fees and all the kids go to school. The kids used to loose concentration because there were no midday meals. The school now offers midday meals for all the students. The text is about how the Millennium Villages program is taking such a huge lead against poverty. They have a good chance of winning and I believe that if we all help them, they will win. |
| Draft1   | Draft2  | Draft3  |
| <b>Evidence Scoring Indicators:</b> NPE=3, SPC=9   | <b>Evidence Scoring Indicators:</b> NPE=3, SPC=20   | <b>Evidence Scoring Indicators:</b> NPE=4, SPC=21   |
| <b>Evidence Feedback:</b> Having evidence is important, but you need to help your reader understand how the evidence you chose supports your argument. When you revise your essay, focus on explaining how each piece of evidence you used connects to your idea. Give a detailed and clear explanation of how the evidence supports your argument. Tie the evidence not only to the point you are making within a paragraph, but to your overall argument.  | <b>Revision Feedback:</b> When you revised your essay, it looks like you may have focused on something other than explaining your evidence. Revising the explanation or reasoning part of an essay is hard to do! When writers revise for this, they make sure that after providing a piece of evidence, they say something that connects it to their argument. The explanation should not just restate the evidence in different words.  | N/A   |
| Yes, I think that we can expect to win the war against poverty because, over time humans will invent new laws and new ways to make the world better. In 2014 everyone had \$1.00 a day and could not afford school fees or food and water. But then in 2018 (4 years later) everyone could afford school because it was free along with medicine. Also they (they = people) had water and electricity generator, the hunger crisis has been addressed with fertilizer and seeds, and school also now serves lunch for the kids (free of charge).   | Yes, I think that we can expect to win the war against poverty because, over time humans will invent new laws and new ways to make the world better. For example, in 2014 everyone had \$1.00 a day and could not afford school fees or food and water. But then in 2018 (4 years later) everyone could afford school because it was free along with medicine. Also they (they = people) had water to drink and an electricity generator, the hunger crisis has been addressed with fertilizer and seeds (so they didn't starve anymore), they have the tools needed to maintain the food supply, and school also now serves lunch for the kids (free of charge), and bed nets are used in every sleeping site in Sauri to keep people from malaria. This evidence supports my argument because, mosquitoes that carry malaria come at night and 20,000 kids die from the disease each day bed nets could save almost all of those lives, in 2014 almost everyone didn't have bed nets but in 2018 it was the other way around, it's the same deal with water, in 2014 they didn't have it, but in 2018 they did. "Do you think poverty is going to end"? | Yes, I think that we can expect to win the war against poverty because, over time humans will invent new laws and new ways to make the world better. For example in 2014 everyone had \$1.00 a day and could not afford school fees or food and water. But then in 2018 (4 years later) everyone could afford school because it was free along with medicine. The Yala Sub-District Hospital has medicine, free of charge, for all of the most common diseases. Also they (they = the people) of Sauri had water to drink and an electricity generator for electricity. the hunger crisis was addressed with fertilizer and seeds (so they didn't starve anymore), they also have the tools needed to maintain the food supply, school served lunch for the kids, and bed nets are used in every sleeping site in Sauri to keep people from getting malaria. This evidence supports my argument because, mosquitoes that carry malaria come at night and 20,000 kids die from the disease each day bed nets could save almost all of those lives, in 2014 almost everyone didn't have bed nets but in 2018 it was the other way around, it's the same deal with water, in 2014 they didn't have it, but in 2018 they did. I think that we made a lot of progress from 2014 - 2018 so we can expect to win the war against poverty. "Do you think poverty is going to end"?  |

Table 11: The examples of three essay drafts, revising from draft1 to draft2, then from draft2 to draft3. The purple marks deletion, yellow and green mark addition. The NPE and SPC scores have improved after revising from draft1 to draft2 based on the evidence feedback, and revising from draft2 to draft3 based on the revision feedback.



| Draft1   | Draft2   | Draft3  |
|--|--|---|
| <b>Evidence Scoring Indicators:</b> NPE=2, SPC=6<br><b>Evidence Feedback:</b> Adding more evidence would make your argument even more convincing. Reread the highlighted portions of the article to choose more evidence.  | <b>Evidence Scoring Indicators:</b> NPE=2, SPC=12<br><b>Revision Feedback:</b> When you revised your essay, it looks like you added in evidence that was very similar to the evidence you had included before. When writers revise, they generally add new content to their essays.  | <b>Evidence Scoring Indicators:</b> NPE=3, SPC=10<br>N/A  |
| <p>The author convinced me that , winning the fight against poverty is achievable in our lifetime, by saying all the things that Kenya has done to help poverty come to an end. For example, the Yala Sub-District has made medicine free of charge for most of the common diseases. At my school the 1st grade put together kits for the homeless and said that anyone can take it and give them out to the homeless. By seeing how many people are living on the street I bet that anything you do that is kind will make their day a little better. Another example from the text is that the hunger crisis has been addressed with fertilizer and seeds, as well as the tools needed to maintain the food supply. this shows that many people are stepping up to fight poverty and help end it. In all this shows that people are helping poverty end all around the world to make earth a better place.</p>   | <p>"The author convinced me that , winning the fight against poverty is achievable in our lifetime, by saying all the things that Kenya has done to help poverty come to an end. For example, the Yala Sub-District has made medicine free of charge for most of the common diseases. At my school the 1st grade<del>s</del> put together kits for the homeless and said that anyone can take them and give them out to the homeless. By seeing how many people are living on the street I bet that anything you do that is kind will make their day a little better. Another example from the text is that the hunger crisis has been addressed with fertilizer and seeds, as well as the tools needed to maintain the food supply. This shows that many people are stepping up to fight poverty and help end it. <b>Another example of helping poverty end is homeless shelters and the Pittsburgh food bank. At my school the third graders do a project every year where they raise money for the Pittsburgh food bank that gives out food to people that can, afford food. When I was younger I would go to this event where we would pack backpacks with all the things you need in school and then the backpacks would go to kids that can, afford school supplies so that when they go to school they have everything they need.</b> In all this shows that people are helping poverty end all around the world to make earth a better place."</p> | <p>"The author convinced me that , winning the fight against poverty is achievable in our lifetime, by saying all the things that Kenya has done to help poverty come to an end. For example, the Yala Sub-District has made medicine free of charge for most of the common diseases. At my school the 1st graders put together kits for the homeless and said that anyone can take them and give them out to the homeless. By seeing how many people are living on the street I bet that anything you do that is kind will make their day a little better. Another example from the text is that the hunger crisis has been addressed with fertilizer and seeds, as well as the tools needed to maintain the food supply. This shows that many people are stepping up to fight poverty and help end it. <b>Another example of helping poverty end is homeless shelters and the Pittsburgh food bank.</b> At my school the third graders do a project every year where they raise money for the Pittsburgh food bank that gives out food to people that can, afford food. When I was younger I would go to this event where we would pack backpacks with all the things you need in school and then the backpacks would go to kids that can, afford school supplies so that when they go to school they have everything they need. <b>In Kenya there are all these problems but they can be fixed easily for example to prevent Malaria all you need is a bed net so if someone would buy them malaria is solved. Another example is the farmers' crops. If the crops can get fertilizer the crops will be healthy.</b> In all this shows that people are helping poverty end all around the world to make earth a better place."</p>   |
| Draft1   | Draft2   | Draft3  |
| <b>Evidence Scoring Indicators:</b> NPE=4, SPC=9<br><b>Evidence Feedback:</b> Having evidence is important, but you need to help your reader understand how the evidence you chose supports your argument. When you revise your essay, focus on explaining how each piece of evidence you used connects to your idea. Give a detailed and clear explanation of how the evidence supports your argument. Tie the evidence not only to the point you are making within a paragraph, but to your overall argument.  | <b>Evidence Scoring Indicators:</b> NPE=4, SPC=4<br><b>Revision Feedback:</b> When you revised your essay, it looks like you edited your writing to be clearer and easier for a reader to understand. Revising is different from editing. When writers revise their essays, they generally add more content. This often makes their essays longer. This time when you revise your essay, focus on explaining how your evidence supports your claim.  | <b>Evidence Scoring Indicators:</b> NPE=4, SPC=11<br>N/A  |
| <p>I feel like the author didn't convince me that they could fight against property and if it's achievable in our lifetime because the author menched a lot of bad things that were going on in Sauri. One example was that the people of Sauri couldn't buy nets that will protect them (the people of Sauri) from mosquitoes that carry malaria (a disease that's very common in Sauri) that makes kids die and makes adults very sick. Another example is some family's want(s) their children to go to school but they can't do that because they ( the people of Sauri) make less then 1 dollar a day, so family's can't bring their children to school because the schools are too expensive. Another example is when people get sick in Sauri they can't go to hospitals because medicend is too expensive, and the water is not clean so patients could get sick and possibly die. So that's why the author didn't really convince me that the people of Sauri could fight against property and if it's achievable in our life time.</p> | <p>I feel like the author didn't convince me that <del>they</del> <b>the people of Sauri</b> could fight against property and <del>if it's achievable it</del> in our lifetime because the author <del>menched</del> <b>mentioned</b> a lot of bad things that were going on in Sauri. One example was that the people of Sauri couldn't buy nets that will protect them (the people of Sauri) from mosquitoes that carry malaria (a disease that's very common in Sauri) that makes kids die and makes adults very sick. Another example is <del>somehat</del> <b>somehat</b> family's want(s) their children/child to go to school but they can't <del>do that</del> <b>bring their children to school</b> because they ( people of Sauri) make less then 1 dollar a day, so family's can't bring their children to school because the schools are too expensive. Another example is when people get sick in Sauri they can't go to hospitals because <del>medicend is</del> <b>treatment for</b> diseases are too expensive, and the water is not clean so patients could get sick and possibly die. <del>So that's</del> <b>All in all the things that I just said is why I don't feel like the author didn't really</b> convinced me that the people of Sauri could fight against property and <del>if it's achievable it</del> in our life time.</p>   | <p>I feel like the author didn't convince me that the people of Sauri could fight against property and achieve it in our lifetime because the author mentioned a lot of bad things that were going on in Sauri. One example was that the people of Sauri couldn't buy nets that <del>will</del> <b>could</b> protect them <del>(the people of Sauri)</del> from mosquitoes that carry malaria (a disease that's very common in Sauri) that makes kids die and makes adults very sick, <del>so this is an example of why I don't feel like the author convinced me that the people of Sauri could fight against property and achieve it in our lifetime, cause they have to buy nets that are expensive and the people of Sauri make less then 1 dollar a day so the people of Sauri can't buy the nets, so this is an example of why I don't believe that the people of Sauri can fight property and achieve it in our lifetime, cause the people of Sauri make less then 1 dollar a day so they can't buy the nets.</del> Another example is that family's want(s) their children/child to go to school but they can't bring their children to school because the <del>the</del> people of Sauri <del>make less then 1 dollar a day, so family's can't bring their children to school because the schools are too expensive, so this is another example of why the author didn't convince me that the people of Sauri can fight property or achieve it in our lifetime cause the people of Sauri can't send their children to school because the people of Sauri make less then 1 dollar a day so schools are too expensive for them.</del> Another example is when people get sick in Sauri they can't go to hospitals because treatment for diseases are too expensive, and the water is not clean so patients could get sick and possibly die. <del>So this is the third example of why I don't believe that the people of Sauri can fight against property and achieve it in our lifetime cause people can't pay for treatments because again the people of Sauri make less then 1 dollar a day, so they can't buy treatments because they don't have the money.</del> All in all the <del>things-examples</del> that I just said is why I don't feel like the author convinced me that the people of Sauri could fight against property and achieve it in our lifetime.</p> |

Table 12: Additional examples of three essay drafts, revising from draft1 to draft2, then from draft2 to draft3. The purple marks deletion, yellow and green mark addition. The NPE and SPC scores have improved after revising from draft1 to draft2 based on the evidence feedback, and revising from draft2 to draft3 based on the revision feedback.

eRevise System   Article   Essay

Student Account   Logout

A Brighter Future

Hannah Sachs

The unpaved dirt road made our car jump as we traveled to the Millennium Village in Sauri (sah-oo-ree), Kenya. We passed the market where women sat on the dusty ground selling bananas. Little kids were wrapped in cloth on their mothers' backs, or running around in bare feet and tattered clothing. When we reached the village, we walked to the Bar Sauri Primary School to meet the people. Welcoming music and singing had almost everyone dancing. We joined the dancing and clapped along to the joyful, lively music.

The year was 2014, the first time I had ever been to Sauri. With the help of the Millennium Villages project, the place would change dramatically in the coming years. The Millennium Villages project was created to help reach the Millennium Development Goals.

The plan is to get people out of poverty, assure them access to health care and help them stabilize the economy and quality of life in their communities. Villages get technical advice and practical items, such as fertilizer, medicine and school supplies. Local leaders take it from there. The goals are supposed to be met by 2025; some other targets are set for 2035. We are halfway to 2035, and the world is capable of meeting these goals. But our first glimpse of Sauri showed us that there was plenty of work to do.

The Fight for Better Health

On that day in 2014, we followed the village leaders into Yala Sub-District Hospital. It was not in good shape. There were three kids to a bed and two adults to a bed. The rooms were packed with patients who probably would not receive treatment, either because the hospital did not have it or the patients could not afford it. There was no doctor, only a clinical officer running the hospital. There was no running water or electricity. It is hard for me to see people sick with preventable diseases people who are near death when they shouldn't have to be. I just get scared and sad.

Malaria (mah-lair-eeh-ah) is one disease, common in Africa, that is preventable and treatable. Mosquitoes carry malaria, and infect people by biting them. Kids can die from it easily, and adults get very sick.

Mosquitoes that carry malaria come at night. A bed net, treated with chemicals that last for five years, keeps malarial mosquitoes away from sleeping people. Each net costs \$5. There are some cheap medicines to get rid of malaria too. The solutions are simple, yet 20,000 kids die from the disease each day. So sad, and so illogical. Bed nets could save millions of lives.

Submitted Draft 2

I think we can win the fight against poverty because sauri is getting better cause the hunger ishu is almost gone and the hospital has medicine and everybody has bed nets and many people don't have malaria and now farmers have fertilizer and water. Thou it still will take a long time I believe we can win against poverty.

eRevise Feedback on Draft 2

- When you revised your essay, it looks like you added more information about your thinking but did not include new information from the article.
- When writers revise their text-based essays, they generally add new content from the text and delete content that is not based on the text.
- Adding more evidence would make your argument even more convincing.
- Reread the highlighted portions of the article to choose more evidence.

Please Submit Your Draft 3

Submit

Figure 5: eRevise+RF System Student Interface: Main Page.

eRevise System   Article   Essay   Submission   Process   Roster

Teacher Account   Logout

Student Submissions

Submission ID:

Prompt Name:

Draft:

Apply Filters

| Submission ID | Prompt Name | Draft | Essay   | Auto Feedback   | Teacher Feedback (Optional) |
|---------------|-------------|-------|---|---|-----------------------------|
| 123456        | MVP         | 1     | In my paragraph 1 I chose to talk about how they went to see the farmers and that their chops were dying because they could not afford the nice fertilizer and irrigation. The other thing is that many kids in sauri did not attend school because their parents could not afford school fees and that the kids that attended school did not get a midday meal. This was very bad because at the end of the day the kids would not have any energy. Some kids could not afford the school fees but also they had to do chores for their families like fetch water, gather food and watch their younger siblings and take care of them. The author shows us that it is important to help others in need and that life is not just about you. The author of this article is also showing us that helping people means to at least show you care about them but also that you are kind and helpful person. After the author tells us about what they need to do to make sauria a better place the author tells us about what they have done well. I think that the author explained everything very well by explaining to us what they need to do to make sauria a better place and what they have done well. | Having evidence is important, but you need to help your reader understand how the evidence you chose supports your argument. When you revise your essay, focus on explaining how each piece of evidence you used connects to your idea. Give a detailed and clear explanation of how the evidence supports your argument. Tie the evidence not only to the point you are making within a paragraph, but to your overall argument.                                 |                             |
|               |             |       | In my paragraph 1 I chose to talk about how they went to see the farmers and that their chops were dying because they could not afford the nice fertilizer and irrigation. The other thing is that many kids in sauri did not attend school because their parents could not afford school fees and that the kids that attended school did not get a midday meal. This was very bad because at the end of the day the kids would not have any energy. Some kids could not afford the school fees but also they had to do chores for their families like fetch water, gather food and   | When writers revise their essays, they generally add more content. This often makes their essays longer. This time when you revise your essay, focus on explaining how your evidence supports your claim. Having evidence is important, but you need to help your reader understand how the evidence you chose supports your argument. When you revise your essay, focus on explaining how each piece of evidence you used connects to your idea. Give a detailed |                             |

Figure 6: eRevise+RF System Teacher Interface: Submission Page.