

Student Use of Automated Essay Evaluation Technology During Revision

Noreen S. Moore[°] and Charles A. MacArthur^{*}

[°] William Paterson University, NJ | USA

^{*} University of Delaware, DE | USA

Abstract: The purpose of this study was to examine how six middle-school students used Automated Essay Evaluation (AEE) technology to revise their writing. Students in a combined 7th and 8th grade Literacy class at one school participated in two in-depth think alouds and semi-structured interviews as they used AEE technology to revise their writing on two separate writing tasks. Constant-comparative analysis of data, including think alouds, semi-structured interviews, and student writing along with a separate quantitative analysis of student revisions revealed themes in three areas: (a) student use of AEE feedback to make revisions; (b) student motivation to revise their writing when using AEE technology; (c) and student understanding and application of AEE feedback during revision. Findings indicated that students who received low scores used AEE feedback to prompt non-surface revisions whereas students with high scores did not. Further, students who used AEE feedback to prompt non-surface revisions made more overall non-surface revisions, revised for different reasons, made more t-unit level revisions, and had more revisions rated as major successes than students who did not use the feedback. Students who used the AEE feedback, MY Editor, were often confused by the grammar and punctuation feedback and had a low success rate using it. However, students were more successful with the spell checker only feedback. In addition, findings show that students were motivated to revise because of the numerical scores the technology assigned their writing. Moreover, knowledge that they would receive a score prompted students to do extensive revising prior to submitting their writing for scoring. Finally, student understanding of the AEE feedback was varied. Implications for classroom use of AEE technology and directions for future research are discussed.

Keywords: writing, revision, automated essay evaluation



Moore, N.S., & MacArthur, C. (2016). Student use of automated essay evaluation technology during revision. *Journal of Writing Research*, 8(x), #-#-. doi: xxx

Contact: Noreen S. Moore, William Paterson University, College of Education, Valley Road, Wayne, New Jersey 07470 | USA – mooren7@wpunj.edu.

Copyright: Earli | This article is published under Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported license.

Automated Essay Evaluation (AEE) is defined as “the process of evaluating and scoring written prose via computer programs” (Shermis & Burstein, 2003). Not only is AEE technology used to numerically evaluate writing quality both holistically and analytically, it is used to evaluate the semantic quality of texts (Landauer, Laham, & Foltz, 2003). AEE systems that incorporate natural language processing capabilities can analyze the discourse structure of writing and provide qualitative feedback on the structure of the writing. AEE systems have been used as purely formative and summative assessment tools, but they have also been packaged as web-based writing instructional programs. Recently, the field has made a shift in terminology from Automated Essay *Scoring* to Automated Essay *Evaluation* systems. The term *evaluation* is now more prevalent given that the capabilities of the technology extend beyond scoring (Shermis, Burstein, & Bursky, 2013). For example, AEE technology can include both quantitative and qualitative feedback for students and additional writing instructional resources for teachers and students such as electronic writing portfolios, prewriting tools, anchor papers/writing models, and sample lesson plans. The current study focuses on the use of AEE as an instructional tool.

One reason to explore the use of AEE in the classroom is that AEE has the potential to offer more feedback and revision opportunities for students than may otherwise be available. Revision can lead to better overall quality texts and provides a space in which writers can learn about evaluation criteria and improve their skills as writers and revisers (Fitzgerald, 1987; MacArthur, 2007; 2012). Another reason to study AEE is that this technology is poised to take on a greater role in K-12 education. Computer-based benchmark and formative writing assessments are being developed by two national assessment consortia: the Smarter Balanced Assessment Consortia (SBAC) and the Partnership for the Assessment of Readiness for College and Careers (PARCC) (U.S. Department of Education, 2012a, 2012b). School administrators looking to adopt this technology need to understand the logic that informed the development of the technology in order to interpret the feedback and scoring provided. They also need to understand how the technology is best used to support writing and revising in the classroom. In fact, Whithaus (2013) contends that AEE technology and other technologies that provide feedback on writing are now a part of students’ writing processes, and it is imperative that writing teachers and researchers participate in active discussion and research on the use and development of these technologies.

Research is just beginning to explore the capabilities of AEE technology as an instructional tool in the classroom. The purpose of the current study was to examine 7th and 8th-grade students’ use of AEE technology to revise their writing. This study was part of a larger study that examined AEE in a classroom context to understand how teachers used AEE to teach writing and to understand student and teacher perceptions of AEE (Moore & MacArthur, 2008).

1. Review of Literature

1.1 Feedback and Revision

A recent meta-analysis on formative assessment and writing (Graham, Hebert, & Harris, 2015) indicated that providing feedback to students about their writing had significant positive effects on their writing quality. One of the reasons that feedback is powerful is that it can prompt revision activity, which in turn can lead to higher quality writing and opportunities to practice and learn about writing (Beach & Friedrich, 2006; MacArthur, 2007; 2012). Quality instructional feedback can help all writers, but in particular struggling writers (Troia, 2006). However, in order for instructional feedback to be useful to students, feedback must target students' learning goals, identify strengths and weaknesses, and provide guidance on improving weaknesses (Hattie & Timperley, 2007). Effective feedback for writing must also help students understand the rhetorical demands of the writing task and help them learn to self-evaluate their writing (Beach & Friedrich, 2006).

Two common sources of feedback are teachers and peers, but both present challenges. Teacher feedback is time consuming, not immediate, and often ineffective (Beach & Friedrich, 2006). Although peers can provide more immediate feedback, peer review requires proper training (MacArthur, 2012; 2015). AEE technology could provide a solution to these time and frequency barriers by giving frequent, immediate feedback to students, which could in turn accelerate the feedback-practice loop which helps writers develop their skills at their own pace (Kellogg & Whiteford, 2009). In addition, due to human error, teacher and peer feedback is often inconsistent; they may not always identify the same problems on a student paper each time they evaluate it. This can send a confusing message to students. In contrast, AEE feedback provides consistent feedback on the same types of errors for every draft a student submits.

1.2 Validity and Reliability of AEE Technology

Much of the research on AEE focuses on the validity and reliability of the technology for assessment (Shermis, Burstein, & Leacock, 2006; Shermis & Burstein, 2013). In terms of reliability, research consistently demonstrates that correlations between AEE and humans are as high as correlations among human raters (Shermis, Burstein, & Leacock, 2006; Shermis & Hamner, 2012). In terms of validity, research indicates that AEE possesses a high degree of construct validity (Keith, 2003) and is related to external measures of writing (Page, 2003). However, opponents of AEE point out that AEE will never be able to measure all constructs of writing as the technology cannot actually read students' writing (Attali, 2013; Ericsson & Haswell, 2006). Researchers have advocated that AEE not be conceived as a replacement for human scoring with the purpose of emulating human scores, but rather as a complement to human scoring that has been proven to be valid and reliable in measuring a subset of the writing construct (Attali, 2013; Deane, 2013).

1.3 Use of AEE Technology in the Classroom

Several studies of student and teacher use of AEE and its effects have been done at the college level or higher (e.g., Kellogg, Whiteford, & Quinlin, 2010; Lai, 2010; Scharber, Dexter, & Riedel, 2008). Less research has been conducted at the K-12 level; this review will discuss studies that examined the use of AEE technology to support teaching and learning at this level. Included in this review are two different types of AEE feedback programs; those powered by Latent Semantic Analysis (LSA) and those developed through model training. Pearson's *Summary Street*, powered by the Knowledge Analysis Technologies engine (KAT) uses LSA for the particular purpose of giving feedback on the adequacy of summary writing in the following areas: content coverage, mechanics, redundancy, and relevancy. This technology is now called *WriteToLearn* and is owned by Pearson. The technologies developed based on model training include Educational Testing Services' *Criterion* powered by the e-rater engine, Vantage Learning's *MY Access!* powered by the Intellimetric engine, and Measurement Incorporated's *Project Essay Grade (PEG)*. *Criterion*, *MY Access!*, and *PEG* provide feedback on writing traits.

Two studies have examined the effects of *Summary Street's* feedback on students' summary writing. Wade-Stein & Kintsch (2004) found that sixth-grade students who received *Summary Street* feedback spent a longer time revising their work and received higher content and quality scores than students who worked without the computerized feedback. Specifically, students who used *Summary Street* feedback had a more balanced coverage of content in the complex articles they read than students who did not have this feedback. A second study compared the effects of *Summary Street* feedback and word processing spelling and grammar feedback on students' overall summary writing (Franzke, Kintsch, Caccamise, Johnson, & Dooley, 2005). Eighth-grade students composed summaries twice a week for four weeks using either *Summary Street* feedback or word processing feedback, depending on the condition to which they were assigned. Results showed that students who used *Summary Street* feedback composed writing that was rated better in quality, organization, content, use of detail, and style than students who used word processing feedback. In addition, students who used *Summary Street* performed better on an independent, comprehension test than students in the control group.

Three additional studies examined the effects of *Criterion*, which provides analytic scores and general feedback on traits of student writing (Attali, 2004; Shermis, Wilson Garvan, & Diao, 2008). One study examined the effects of AEE and practice on the overall writing score and error reductions by students in grades 6-8 and 10 (Shermis, et. al., 2008). The students wrote to seven writing prompts across the course of a year and received AEE feedback. In general, the results indicated that students improved their overall AEE quality scores, essay length, and number of unique words. Additionally, for most error codes used in the analysis, student errors were reduced over time. In a second study, Attali (2004) analyzed a set of 30,000 essays submitted to *Criterion*. Of the essays that were revised, average improvements were noted on the traits that were

evaluated by the system. Both of these studies had the same major limitations: the researchers did not control for practice effects, and the studies used *Criterion's* own scores rather than independent measures of quality and errors.

Finally, a more recent study examined student use of *Project Essay Grade's* (PEG) feedback to revise their writing (Wilson, Olinghouse, & Andrada, 2014). The study showed that students in grades 4-8 who used the automated feedback in the program produced writing that improved in quality across revisions. However, students appeared to benefit only so much from similar feedback on the same essay; there appeared to be a ceiling effect. In addition, although essays improved in quality from draft to draft, essays on new topics were not better.

In summary. Several studies in this review showed that when students are using the technology the quality of their writing improves, but some studies had major limitations which preclude us from drawing a sound conclusion about the effects of AEE on writing quality. Further, one study showed that once students wrote outside the system, there were no significant improvements. These findings beg the question: what are students learning about writing through practice with AEE feedback? In addition, no research that includes an in-depth view of how teachers and students interact with this technology in the classroom is available. In fact, to our knowledge, there are no studies that have actually collected data on students using the technologies in real time in the classroom at the K-12 level. Instead, researchers have examined data collected by the systems in an effort to understand how the technologies are being utilized and how they impact student writing quality rather than observational, interview, and think aloud data of students using the AEE technology. Studying students' writing and revision processes with AEE technology could offer new insights on previous research findings as well as provide a more nuanced picture of how students use the technology.

Specifically, we need to know what kinds of revisions students make using AEE and whether the revisions are successful. Research has shown that students spend a longer time revising when using AEE. However, we need to understand what motivates students. Finally, research illustrates that even though students improve the quality of their writing while working with AEE, their learning does not appear to transfer. Therefore, it is important to delve deeper to understand what students are learning through their use of AEE technologies. We need to explore how well students understand the feedback from AEE, both the substantive feedback and the conventions feedback, and how they use the feedback to make revisions. Answering these questions demands that researchers look beyond the data collected by the AEE technologies themselves; instead, classroom observations, interviews, think-alouds, and video of student use of the technology need to be studied in a systematic way.

1.4 The Present Study and the Context

The current study is part of a larger case study that aimed to address gaps in research on the use of AEE technology to support writing development and writing instruction in the

classroom (Moore & MacArthur, 2008). The purpose of the larger study was threefold: (1) to understand how teachers in one 5th and one 8th grade classroom at two different schools used AEE technology to support writing development and instruction with a specific emphasis on revision; (2) to understand how teachers and students perceive AEE technology; (3) to understand how students use AEE technology to support their learning. The current study focuses on how students used AEE technology to support their writing and revision. The research questions that guided this work are:

- a) How do students use AEE feedback to make revisions?
- b) Are students motivated to make revisions while using AEE technology?
- c) How well do students understand the feedback from AEE, both the substantive feedback and the conventions feedback?

The current study extends and contributes to previous research in two ways. First, we examine in detail the types of revisions made by students. Second, we examine students' use of the program through in-depth think alouds and semi-structured interviews in order to get a closer look at how students interact with AEE programs and feedback. Third, we focus our data collection on students in 7th-8th grade in order to contribute to research on K-12 usage of AEE technology to support writing.

2. Method

2.1 Participants

Six students in a combined seventh- and eighth-grade literacy class at a private school in the Northeast part of the United States participated in the study. The school they attended served approximately 60 students in grades one through eight. All students demonstrated learning, attention, mild social/emotional and/or mild behavioral issues that may interfere with school success. The sixth, seventh, and eighth grade literacy classes used AEE as a supplement to their writing instruction which included topics on genre, word choice, peer revision, and the process approach. Data collection took place during the third year the school was using AEE. It was also the third year the teacher was using AEE and the second year the students were using AEE. The school had a 1:1 computer to student ratio; all students worked on AEE in their classrooms on a regular basis during their literacy block.

This classroom was comprised of twelve students (5 female; 7 male) ranging in age from 12 to 14 (Table 1). Four students were performing at or above grade level and eight students were performing below grade level. The primary language of all students was English; 17% of the students were black and 83% were white. In conjunction with the classroom teacher, we selected six students who represented a range of abilities and genders to participate in the study.

Table 1: Classroom Demographics

	Classroom	Target Students
Number of Students	12	6
Gender		
Male	7	3
Female	5	3
Age Range	12-14	12-14
Race/Ethnicity		
Black	17%	16.7%
White	83%	83.3%
Hispanic	0%	0%
Primary Language		
English	100%	100%
Other	0%	0%
Months using AEE	5-16.5 months	16.5 months

2.2 Automated Essay Evaluation (AEE) Technology

The school used the 2008 version of the web-based MY Access! automated essay evaluation and instructional system (Vantage Learning, 2008). Since 2008 the following aspects of the program have been updated: improvements in usability, audio MY Tutor feedback, writing tasks linked to the Common Core, multimedia support, and a non-flash based writing environment (Vantage, personal communication, June 19, 2014). Intellimetric is the intelligent scoring system that underlies the MY Access! system (Schultz, 2013). It uses artificial intelligence (AI), natural language processing (NLP), and statistical technologies to produce scoring models used to assess student writing (Elliot, 2003; Schultz, 2013). Pearson correlations between human raters and Intellimetric for holistic scores by grade level are: .93 (elementary), .92 (middle school), .91 (high school), .83 (higher education) (Schultz, 2013). Agreement within one point was reported as: 100% (elementary), 99% (middle school), 99% (high school), and 90% (higher education) (Schultz, 2013). No information on the reliability of the trait scores has been reported. In terms of face validity, Intellimetric is trained to analyze semantic, syntactic, and discourse-level features associated with writing quality: focus/coherence, organization, elaboration/development, sentence structure, and mechanics/conventions.

When students log on to the program they are assigned an existing prompt (narrative, informative, persuasive, text-based, or literary) or a teacher-created prompt. Students can compose using the word processor in the program. There are prewriting (e.g., graphic organizers) and feedback tools available to students as they plan, write, and revise.

As they write, students can ask MY Editor for feedback on spelling, grammar, and punctuation, or they can ask MY Tutor for more global feedback on the traits on which

their writing will be evaluated: focus and meaning, content and development, organization, language use and style, and conventions. MY Tutor feedback is qualitative feedback tied to the trait, score, and genre. MY Tutor feedback has three parts: 1) a brief one sentence revision goal (e.g., make your characters more realistic), 2) strategies for achieving the goal (e.g., highlight the details you included about your characters, now add more); 3) and a before and after example of a student revising based on the revision goal and strategy.

Once students submit their essays for evaluation, MY Access! generates an immediate quantitative and qualitative feedback report to help them improve their writing. Students receive either a holistic score and/or analytic scores depending on the setting the teacher selects. Students receive two types of qualitative feedback in their score report: MY Tutor and MY Editor. Although the MY Tutor feedback is different for each score point and genre, the same feedback is given for the same score in the same genre. MY Editor feedback is in-line feedback on grammar, punctuation, and spelling and is specific to each student's writing. Individual errors that students make are highlighted and accompanied by a specific comment about what could be wrong and suggestions for how to fix the error. Students can also opt to use only the spell checker feedback. For each writing prompt, all students' drafts with scores and feedback are stored in an online portfolio that can be accessed at any time.

2.3 Data Collection

Data collection occurred over the course of a two-and-a-half-month period at the school. The primary researcher collected twelve think alouds, twelve semi-structured interviews, and 24 drafts of student writing on two writing tasks: task one was a narrative task ("To Change a Day in Your Life") and task two was an argumentative writing task ("Year-Round Schooling"). Think Alouds were collected using free software, CamStudio, which recorded both audio and screen images so we could hear students think aloud and watch what they were doing on the screen simultaneously. Semi-structured interviews were audio recorded. Student writing was copied and pasted from the AEE technology into MS Word and saved in a protected file on the researcher's computer. All data sources associated with each student were grouped and organized into separate folders. Multiple, complementary data sources were collected in order to ensure triangulation of the data. Additional detail about think-alouds, semi-structured interviews, and student writing are provided in Appendices A, B, C.

2.4 Data Analysis

Analysis took place throughout data collection, as the primary researcher transcribed, organized, read and wrote memos as data was collected. Constant-comparative analysis (Strauss & Corbin, 1998) was used as we identified categories that were related to our research questions and which captured the participants' stories. Think aloud data was transcribed using guidelines established by Prior (2003) to distinguish between writing, thinking, rereading what the writer wrote, oral composition, reading another

text, rereading another text, or some other action. Semi-structured interview data was also transcribed. Finally, before and after drafts of student writing were compared using the compare feature in MS Word to identify all revisions. Papers were also cross-referenced with think aloud transcripts to ensure that all surface and non-surface level revisions were identified.

Constant comparative analysis of think-alouds, semi-structured interviews, and writing.

The researchers read through all cases, applied codes and wrote memos which were discussed and revised repeatedly. Triangulation among multiple data sources as well as ongoing researcher discussions were used to ensure validity. The researchers identified three main themes: (a) revisions made by students working with AEE; (b) student motivation to revise their writing when using AEE technology; and (c) student understanding and application of AEE feedback during revision.

Non-surface and surface-level revision analysis.

Drafts written before and during the think aloud sessions were analyzed for individual revisions. All changes made to a text during a think aloud session *after* students submitted their draft and received scores and feedback were counted as revisions. If students made changes to a draft prior to the think aloud or during a think aloud session, but prior to submitting their draft for initial scoring and feedback, these changes were noted, but not included in the revision analysis. Transcriptions of the think alouds and the think aloud videos were looked at simultaneously to help code the revisions.

Revisions were coded using a scheme modified from previous research (MacArthur, Graham, & Schwartz, 1991). Surface revisions include spelling, punctuation, capitalization, and grammar. Surface revisions were coded for the following: type, source prompting revision, and quality. Non-surface revisions included all other revisions. Non-surface revisions were coded for the following: (1) level (i.e., word, phrase, t-unit); (2) operation (i.e., add, delete, substitute); (3) reason for revision (i.e. to add detail, to focus writing, to clarify, to engage audience, to transition, to make a complete sentence); (4) source prompting revision (i.e., AEE feedback, something a peer said, something a teacher said, something the student thought of while rereading); (5) and quality (i.e., major improvement, minor improvement, or no change). See Appendix D for definitions of the revision codes.

Both researchers coded the revisions independently. Interrater reliability (agreements/[agreements + disagreements]) was 94.9% for non-surface revisions and 96% for surface level revisions. The reliabilities for individual categories of the non-surface and surface revisions were all over 90% exact agreement. All disagreements were resolved through discussion.

3. Results

The results section is structured around three central themes: student non-surface and surface revisions, student motivation, and student understanding and application of AEE feedback. Each theme includes examples illustrating them in more depth.

3.1 Student Non-Surface and Surface Revisions

To answer our first research question, about student use of AEE feedback to make revisions, we conducted a quantitative analysis of non-surface and surface revisions in conjunction with the qualitative analysis of think aloud and semi-structured interview data.

Overall, we found that students used MY Tutor feedback to prompt non-surface revisions only when they received scores they perceived as low. Three students, Alex, Gabriella and Meredith, received scores below the maximum and used MY Tutor feedback to prompt non-surface revisions (see Table 2).

Table 2: Revisions after receiving scores by task

Writing Task	Alex		AJ		JD		Kara		Gabriella		Meredith	
	1	2	1	2	1	2	1	2	1	2	1	2
Scores Prior to Revising	5.2	6.0	6.0	6.0	5.9	6.0	6.0	6.0	5.1	4.9	5.3	5.8
Used MY Tutor	Y	N	N	N	N	N	N	N	Y	Y	Y	Y
Used MY Editor	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N
Used Spellcheck Only	N	N	N	N	N	N	N	Y	N	N	Y	Y
# Non-Surface Revisions	8	0	0	10	0	0	0	3	11	6	11	8
# Surface Revisions	7	2	0	15	1	1	0	1	3	0	3	2

Alex used MY Tutor feedback on task one whereas Gabriella and Meredith used it on both tasks. Further, in these three cases, students did more non-surface revising than surface-level revising. In contrast, three students, Kara, AJ, and JD received scores at or near the maximum and did not use the MY Tutor feedback or make many non-surface revisions.

There was one outlier in the pattern; on task two, AJ made 15 non-surface revisions after receiving a top holistic score of 6. However, he did not use MY Tutor; his non-surface revisions were prompted by MY Editor feedback. When using the MY Editor tool, AJ was often confused by the advice it was giving and this prompted him to rewrite or add phrases or words in order to clarify his writing and to make the MY

Editor flags disappear. This may be why feedback designed to prompt edits to grammar, punctuation and spelling, actually prompted larger level, non-surface revisions. Finally, it was clear throughout the think-alouds and semi-structured interviews that students used MY Editor when they didn't feel they had much more revising to do or when they received top scores.

Non-surface revisions

Five out of six students made at least one non-surface revision during the course of the study, but there is variation across students in terms of total revisions, level, operation, reason for revision, success rate, and source prompting the edit (See Table 3). The students who used MY Tutor feedback during revision (i.e., Meredith, Gabriella, and Alex) made the most non-surface revisions, made more revisions at the t-unit level, revised for a wider variety of interesting reasons, and had a higher rate of major success than the students who did not use this feedback.

For both Gabriella and Meredith, who made the most revisions, there is a clear distinction between the success rate, level, and the purpose of revisions prompted by MY Tutor and those prompted by other sources. Revisions prompted by MY Tutor had more instances of major success, more instances of t-unit level revisions, and included the following reasons for revision: audience engagement, focusing, adding argumentative elements, and transitioning. For example, Meredith made 12 revisions prompted by MY Tutor with purposes to improve transitions, focus the paper, and engage the audience, and add argumentative elements (reasons, counterarguments), whereas the six revisions prompted by self-evaluation and the teacher mostly focused on adding details. Further, her revisions prompted by MY Tutor were longer and more successful. Seven of her revisions prompted by MY Tutor were rated as major successes, four were rated minor success, and only one was rated no success.

Three students did not apply MY Tutor feedback to make non-surface revisions. Kara was prompted by self-evaluation only, AJ was prompted by both self-evaluation and MY Editor, and JD did not make any non-surface level revisions. Kara made three non-surface level revisions and she made these revisions for the purpose of clarifying and adding details. Two of her revisions were rated as minor successes and one was rated unsuccessful. Two were made at the phrase level and one at the word level.

AJ made a total of 10 non-surface revisions. Most were at the word and phrase level and rated as minor successes; all were done for the purpose of clarifying. As previously stated, nearly all of these revisions were prompted by the MY Editor tool, when he was confused by the editing advice and simply rewrote or added words to clarify his writing and to make the MY Editor flags disappear. Finally, although JD did read MY Tutor, he did not make any non-surface revisions.

Table 3: Non-Surface Revisions

	Totals	Alex	AJ	JD	Kara	Gabriella	Meredith
Total Non-Surface Revisions	57	8	10	0	3	17	19
Level:							
T-unit	23	8	1	0	0	4	10
Phrase	21	0	7	0	2	6	6
Word	13	0	2	0	1	7	3
Operation:							
Add	25	6	1	0	1	6	11
Delete	8	1	1	0	1	4	1
Rewrite	24	1	8	0	1	7	7
Success:							
None	8	0	2	0	1	4	1
Minor Local	32	2	8	0	2	9	11
Major	17	6	0	0	0	4	7
Reason given for Revision:							
Add detail	13	6	0	0	0	2	5
Improve focus	6	2	0	0	0	3	1
Engage Aud.	5	0	0	0	0	0	5
Clarify	19	0	10	0	2	5	2
VaryWord	7	0	0	0	1	3	3
Comp. Sent.	3	0	0	0	0	1	2
Transition	4	0	0	0	0	3	1
Source Prompting Revision:							
Self	20	0	1	0	3	10	6
Teacher	1	0	0	0	0	0	1
MY Tutor	27	8	0	0	0	7	12
MY Editor	9	0	9	0	0	0	0

Surface revisions

All students made at least one edit during the course of the study, but there was variation in total edits, type, success rate, and source prompting the edit (See Table 4).

Students who used MY Editor feedback made the most surface revisions; however, they experienced a low success rate. Students who used self-evaluation or the Spellcheck only had a greater success rate.

Table 4: Surface Revisions

	Total	Alex	AJ	JD	Kara	Gabriella	Meredith
Total Edits	35	9	15	2	1	3	5
Type:							
Spelling	12	3	3	0	1	0	5
Punctuation	7	1	4	0	0	2	0
Grammar	16	5	8	2	0	1	0
Success:							
No Success	19	5	10	2	0	1	1
Success	16	4	5	0	1	2	4
Source:							
Self	10	5	2	0	0	3	0
MY Editor	19	4	13	2	0	0	0
Spellcheck	6	0	0	0	1	0	5

For example, AJ and Alex used the MY Editor feedback the most and also made the most surface level revisions. AJ made 15 edits on task two, using MY Editor to prompt 87% of them. A little over half of AJ's edits focused on word-level grammar, but he also made spelling and punctuation edits. Most (67%) of his total edits and 90% of those prompted by My Editor were rated unsuccessful. Similarly, Alex made 9 edits, 44% prompted by My Editor, and 56% unsuccessful. In contrast, Kara and Meredith only used the Spellcheck, and nearly all of their edits were successful.

Overall, the students focused mainly on word-level grammar edits and spelling and gave less attention to punctuation. Nineteen (71.4%) of the edits were prompted by AEE feedback. Interestingly, over half (54.3%) of the total edits made were rated unsuccessful. Further, MY Editor prompted 68% of the unsuccessful edits. When looking specifically at AEE tools, MY Editor resulted in 8 successful edits and 11 unsuccessful edits. Spellcheck resulted in 5 successful edits and 1 unsuccessful edit.

3.2 Student Motivation

To answer our second research question, about student motivation to make revisions while using AEE technology, we drew from our qualitative analysis of think-alouds and semi-structured interviews. Overall, our data analysis showed that all students were motivated to revise because of the numerical score feedback the technology assigned

their writing. After submitting their writing and seeing their scores, all students' expressed emotions such as "wow," "whoa" "ugh...that's pretty bad." Scores provided a gauge to let students know how they were doing; they gave students a sense of satisfaction with their work or a feeling that they needed to keep working. Scores prompted students to set goals for revision based upon their lowest scores or signaled it was time to stop revising.

Student motivation and non-surface revisions after receiving scores

Think aloud data and semi-structured interview data confirm that low scores prompted students to want to revise their writing at the non-surface level whereas high scores did not.

Gabriella, Meredith and Alex each received scores they perceived as low on at least one task and as a result were motivated to revise. For example, Gabriella was one of the low scorers in the group. Gabriella received holistic scores of 5.1 and 4.9 on task one and two respectively. After getting her scores, she made a total of 14 non-surface revisions on task one and 6 revisions on task two. When she saw her scores on task two, she exclaimed, "Wow. That's pretty bad. OK." She explained that she wanted to "get [all] 5's or if it's possible 6's." The scores prompted her to read the MY Tutor feedback and revise during the think-aloud session. Further, when I went back into the AEE system to look at her portfolio several days after the think aloud, I found that she continued revising her writing even after our meeting and eventually raised her analytic and holistic scores to 5s on task two.

In contrast, AJ, Kara and JD received high scores on at least one task and showed low motivation to revise on these tasks. JD received high scores on both tasks and also did the least revising. He received holistic scores of 5.9 and 6.0 on task one and two respectively and only made one surface level change to each paper. During his think aloud for task two, he exclaimed, "Wow! Wooooow!" and whistled in delight. During the rest of the think aloud, JD did not revise and did very little editing. He mostly read and clicked through the MY Editor feedback, but rarely applied it. Although he didn't explicitly state that he felt he was finished, when I probed him to tell me what his next moves were with his writing, he said he would probably "just sit there and I don't know."

Motivation to revise prior to receiving scores and feedback

We noticed an interesting trend regarding revision that happened prior to students' receiving AEE feedback and scores. Knowledge that they would be scored seemed to drive students to write and revise extensively prior to receiving feedback.

During the think-aloud and semi-structured interview, five students stated that they always spent time rereading and revising their writing before they submitted their writing for scoring. Three students, AJ, Kara, and Gabriella, also demonstrated this tendency during their think alouds. The think aloud protocol used in this study asked students to log on and submit their writing for scoring and feedback prior to revising.

However, when AJ, Kara, and Gabriella met with me for the think aloud, they ignored my direction to submit their writing first and instead, they reread, revised, and edited until they felt their writing was “correct” or their “best work” before submitting it for evaluation.

For example, after giving Kara the think aloud directions, she clicked the “initial submit” button, but doing so takes users to a read-only review page where they can read through their writing one more time before clicking “final submit.” Kara reread her writing on the review page, noting, “I’m reviewing it to see if there are any corrections.” She found something she wanted to add in the beginning of her narrative piece and she clicked the back button to return to the main writing screen. She revised for 35 minutes and made a total of eight revisions before submitting her writing for AEE feedback and scores. In her semi-structured interview, I probed Kara about how she typically decides when to submit her writing. She explained, “I look my essays over three times to make sure I didn’t miss anything like spelling errors, mechanic errors and language use, then I submit it and see what I got.”

Only one student, JD, reported he did not have the need to submit only very polished writing. At the end of the think aloud task two, JD mentioned that he could “submit as many times as [he] want[s]” so he didn’t consider it his “final submit” when he clicked the button. He also explained that if he did get low scores, “[he]’d probably submit it again and see if [his] scores went up and if they didn’t [he]’d keep working and keep kind of submitting it every time [he] made a change or a couple of changes.” JD’s story checks out when looking at his writing portfolio saved in the AEE system. When he met with me to do the think aloud for his persuasive essay, he had already submitted a total of 9 drafts and there were very minor changes between each draft (e.g. one spelling correction, addition of one comma, etc.).

3.3 Student Understanding and Application of AEE

In this section, we will discuss findings regarding our third research question, how well students understood the feedback from AEE, both the substantive and conventions feedback. There was variation across students regarding understanding of MY Tutor. However, all students who used MY Editor appeared to have difficulty understanding this feedback.

Student understanding and application of MY Tutor Feedback

Students could be grouped into two categories with subcategories based upon understanding and use of MY Tutor Feedback: (a) three students read and applied the feedback; two of those students, Alex and Meredith, demonstrated a good understanding and application of the feedback and one student, Gabriella, demonstrated limited understanding and application of the feedback; (b) three students did not apply the feedback; two students, AJ and Kara, read the feedback, but did not apply it for different reasons and one student, JD, did not read or apply the feedback.

Good understanding and application

After she submitted a first draft of task one, Meredith was visibly unhappy with her scores. In her think aloud, she said, “I think I really need to work on organization and content.” Meredith skimmed through the MY Tutor feedback to find what it said about these trait areas only. She didn’t read the feedback straight through in its entirety; instead she stopped on bold-faced goal statements under each trait area. She also occasionally read the commentary written to accompany a before-and-after example of a revision, which was given in the feedback. For her narrative piece, she first stopped to read the commentary under the feedback for Content and Development. She read the feedback which included a commentary by a fictional writer who revised a small section of his writing to include more details. The feedback shows a before and after example of writing and includes the writer’s thoughts on how his revision improved his paper. She thinks aloud, “[he wanted] to make sure all of [his] details related to [his] story.” Then she thought aloud, “So, I guess I need to relate more things to the story instead of going off topic.” Next, she scrolled down to the next feedback suggestion in bold and read, “Add dialogue.” She thought aloud, “Ok, so I’ll add dialogue and what else was there to do?” She scrolled down to the organization feedback and read a goal in bold face, “Use transitions (words such as morning, before, now, suddenly, yesterday)” to show how events in your story go together.” She thought aloud, “Alright so I guess my plan is to go back and change, add some dialogue and some detail and also that’s for content and for organization I will go back and add a few transitions...better transitions. During Meredith’s think aloud she added dialogue and transitions to her narrative.

Limited understanding and application

Gabriella was also visibly disappointed when she received the AEE evaluation on task two. In her think aloud, she exclaimed, “Wow, that’s pretty bad!” after seeing her scores. She decided to look at the MY Tutor feedback to help her revise her organization, which was her lowest score. The first feedback tip Gabriella read suggested that she reread her essay to ensure that each of her reasons for her opinion is in its own paragraph. In addition, the feedback suggests that she consider the order of importance for her reasons. After reading this feedback, Gabriella says, “So, I’m thinking I could make the beginning paragraph more catching, because I said “first, second, and third, and that’s kind of boring so I’m thinking I could do something different.” Clearly, Gabriella’s think aloud demonstrates a lack of understanding of what the feedback is asking her to do. She also reads a tip that suggests she ensure that she has transitional words and phrases to connect her ideas. As she reads, she has difficulty decoding the word “transitions” and instead reads the word transition as translation. After she is finished reading this feedback, she says, “So, I’m thinking maybe it shouldn’t be fact after fact. I’m thinking to make it more strong, I could take sentences from...take...make...more sentences from one fact that I picked or I could...yeah...so.” In this excerpt from the think aloud it is also clear, that Gabriella lacks a full

understanding of what the feedback is suggesting she do. Thus, it is not surprising that when Gabriella went back to her paper to revise she did not apply the feedback. She decided to rewrite the transition words she already had in her first paragraph. Originally, she structured the paragraph to include an opinion statement and three reasons. She connected the reasons with the following transition words: one reason, my second thing is, and lastly. During revision, she rewrote those transition words and phrases to be the following: first, my second thought is, and my last reason is."

Despite her limited understanding in this example, Gabriella did apply the feedback successfully to engage the audience in task 2. For example, she read MY Tutor feedback about including a strong call to action to cause her audience to take a stand. She adds the sentence, "So please think about the choice you are making" to the end of her piece. She explains that she was trying to "make people realize how important" this cause was. In task 1, Gabriella read feedback about narrowing her focus. When revising she deleted an entire sentence. She thought aloud as she did, "I'm thinking about how they said narrow focus. I'm thinking about maybe I can take out that part about candy because it doesn't ...it has something to do with her [main character] in the hospital...so I think that like I could delete that part."

No application

Three students, Kara, AJ and JD, did not attempt to apply the MY Tutor Feedback. Whereas AJ and Kara actually read some of the MY Tutor feedback, but chose not to apply it, JD did not read it at all. AJ demonstrated a good understanding of the feedback he read, but did not feel it was necessary to apply it since he already had high scores. On the other hand, Kara had some difficulty understanding the feedback and how to use it. The main source of her lack of understanding stemmed from her confusion about what procedures the feedback was asking her to follow. That is, one part of the MY Tutor feedback gave writers a process for highlighting and evaluating aspects of their writing before making changes (e.g., Highlight your topic sentences. Do you have a topic sentence in each paragraph? If not, add one). After reading feedback that asked her to highlight her reasons, Kara turned to me and said, "Um...it says for me to highlight...should I?" As she read the feedback she never thought aloud about goals or how it related to her writing. When she finished reading the feedback she said, "I might highlight and underline the things they said and um look over it and see if I have any spelling errors." When she went back to her own writing, she did not highlight and underline, she reverted back to her own strategies for revising, which included rereading and self-assessment.

Student understanding and application of AEE MY Editor and Spelling feedback

The three students who used MY Editor grammar and spelling AEE feedback, AJ, Alex, and JD, experienced difficulty understanding and applying it. Lack of understanding, confusion, and frustration stemmed from unfamiliarity with the technical vocabulary

used or from the fact that the feedback was incorrect. As a result, students guessed ways to correct their grammar, punctuation, and spelling mistakes when words were flagged or they simply ignored the feedback.

During task 2, AJ made extensive use of MY Editor, but was frequently frustrated by the feedback. For example, MY Editor flagged a “clause error” in the following sentence: “Added on to that should be a reading packet, a writing packet to review mechanics and other writing errors, and a book report that is due on the first day of school.” The error itself is difficult to discern; we think the technology was referring to a vague use of the word “that.” In the example below, AJ tried to address the flagged error, but it is unclear he understood what was incorrect in the first place and what he needed to do to fix it. While editing, he actually introduced a new, but minor error by adding a comma after “mechanics.” He also introduced a logical error by stating a book report should be added to the book report. He thinks aloud:

Yeah, I know what they want. They want me to take out the...Or, no...They want me to um...[AJ changes the sentence to “Added on to the book report, there should be a reading packet, a writing packet to review mechanics, and other writing errors, and a book report that is due on the first day of school.] And maybe that will be right? [AJ clicks MY Editor to see if error flag was removed]. Yep!

Students could become confused even when they eventually realized that a suggestion was incorrect. In the following example, MY Editor incorrectly suggested that JD replace “grade” in the following sentence with “rank”: “If a child misses 12 days of school in certain states he or she will have to repeat that grade.” JD is confused by this suggestion and he thinks aloud about it:

Even though I can change a word... and it could be a synonym it won't make sense...like I put “grade” and it says I should change it to “rank,” but if I change it to rank it's gonna sound really weird.

In many of the instances when students disagreed with MY Editor or were confused by the feedback, the feedback seemed to be incorrect. It appeared that the AEE program was misinterpreting the syntax or word usage, and as a result gave advice that was technically wrong. However, a more in-depth analysis would need to be conducted to confirm the frequency with which this occurs in the program. Still, this observation is not out of line with previous research. For example, a recent analysis of the feedback in MY Access! also demonstrates that the MY Editor! feedback was often incorrect or failed to flag errors (Dikli, 2010). In addition, MacArthur (2000) reviewed grammar checkers and found that these tools were likely to miss errors in student writing or incorrectly flag errors in student writing.

4. Discussion and Implications

AEE programs as instructional tools are being packaged and sold to schools for classroom use. However, there is a serious lack of research on how to best implement and use these tools to promote student learning. This study takes one of the initial steps in providing insights in this area. Overall, students in this study appeared motivated to revise by receiving scores from the AEE technology. Further, students used substantive MY Tutor feedback to revise when they received low scores; they used conventions MY Editor and Spellcheck feedback when they felt they had nothing substantial to revise. When students used the substantive AEE feedback they made more revisions, revised at the t-unit level, revised for different purposes, and were more successful than students who did not use the feedback. However, students' understanding of this substantive feedback was varied. When students used the MY Editor conventions feedback they were often confused by it, and it often didn't result in successful edits. On the other hand, students used the Spellcheck feedback with greater ease and success.

4.1 Limitations of the Study

The results of this study should be interpreted with three limitations in mind. First, because a very small, unique sample was used, it is impossible to make generalizations to other contexts where AEE technology is being used. On the other hand, a study of this size allowed us to gather rich data and provide detailed description of what student use of AEE looks like. A larger sample may not have allowed such a fine-grained analysis.

Secondly, the students in the study worked with one AEE technology. It is possible that other technologies and the feedback given would influence student use in different ways. Further, it is possible that other technologies have different features students could use and which were not explored in this study.

Finally, as the principal researchers, the methodological choices we made influenced the research. For example, we defined revisions as all changes made during the think alouds, but we learned that students revised outside of this window. We did not capture and analyze revisions outside of the think aloud in great detail.

4.2 Implications for Instruction

Our analyses suggest several areas teachers, coaches, and instructional leaders who are using or considering use of this technology should bear in mind. When students received scores they perceived as low, they used the substantive MY Tutor feedback and improved their writing. This is promising because it shows the AEE technology can encourage students to revise and help them improve their writing. However, in this study, some students received high scores on their first submissions, which led them away from using MY Tutor feedback because they felt they did not need to revise. Scoring models in AEE technology need to match individual classroom contexts. Teachers, coaches, and leaders need to know how to call upon the use of more difficult scoring

models within AEE systems (i.e., different grade level models within the system) when students begin to consistently score at the top of the scale.

Further, students in this study did not always understand the MY Tutor substantive feedback. Students require instruction and teacher modeling on how to use the AEE feedback and evaluation rubric. Such alignment between instruction and technology would better allow students to take full advantage of the feedback offered in the system and improve their writing. When students learn evaluation criteria, they can apply that knowledge to improve their writing (MacArthur, 2012). AEE technology can remind them of the criteria they learned and will provide an additional outside assessment of their ability to meet the criteria in the rubric.

The students had difficulty understanding MY Editor feedback, and this resulted in a low rate of successful edits prompted by the technology. Students require instruction on the grammar and punctuation errors being flagged in the system so they understand the vocabulary used to talk about the error and how to address the issue. Again, this alignment between classroom instruction and technology would allow students to take full advantage of the technology to improve their writing. That being said, we cannot expect the technology to be correct in every instance. Therefore, students need to be taught strategies for evaluating and deciding whether to use or ignore the feedback given.

4.3 Future Directions for Research

In this study, scores prompted revisions both before and after students submitted their writing for a score. It is important to follow up on this trend to understand if students spend more time revising when using AEE as compared to writing outside of this technology. Further, it is important to examine what specifically students were learning from the technology (i.e., what earned them a high score) and how this impacted their beliefs about writing.

Another future direction is to examine the impact of different AEE feedback on student learning more closely. We found there was variation in terms of student understanding of AEE feedback in this system. It is important to gain a better understanding of the different characteristics of feedback offered by different systems. We need to know which characteristics of effective feedback in AEE technologies are effective for individual students.

Finally, the research reported in this paper did not offer a picture of how and what teachers did to implement AEE. It is important to understand how teachers perceive and use AEE technology as well.

In conclusion, as AEE technology finds its way into the writing classroom, it is imperative that we understand how to best use it to support teaching and learning. The technology holds promise as a tool for promoting revision and helping students develop evaluation criteria to be applied during revision. This in turn will help students improve their writing and improve their abilities as writers. However, the people developing the technology and the teachers, coaches and leaders using the technology must

collaborate in order to develop a tool that teachers and students can use and which aligns to classroom instruction. Moreover, teachers need an understanding of how to best implement the technology so that it can enhance rather than replace or detract from the instruction they already have in place.

References

- Attali, Y. (2004). *Exploring the feedback and revision features of Criterion*. Paper presented at the National Council on Measurement in Education. San Diego, CA April 12-16.
- Attali, Y. (2013). Validity and reliability of automated essay score. In M.D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 181-198). New York: Routledge.
- Bardine, B., Bardine, B., & Deegan, E. (2000). Beyond the red pen: Clarifying our role in the response process. *English Journal*, 90(1), 94-101.
- Beach, R., & Friedrich, T. (2006). Response to writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 222-234). New York, NY: Guilford.
- Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing* 18: 7-24.
- Dikli, S. (2010). The nature of automated essay scoring feedback. *CALICO Journal*, 28(1), 99-134.
- Elliot, S.(2003). IntelliMetric: From here to validity. In M.D.Shermis & J.Burstein (Eds.), *Automated essay scoring: Across-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Erlbaum Associates.
- Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research*, 57, 481-506.
- Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., Dooley, S. (2005). Summary Street®: Computer support for comprehension and writing. *Journal of Educational Computing Research*, 33(1), 53-80.
- Graham, S., Hebert, M., Harris, K.R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, 115(4), 523-547.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Keith, T. Z. (2003). Validity and automated essay scoring systems. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 147- 168). Mahwah, NJ: Erlbaum.
- Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist*, 44, 250-266.
- Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students to write? *Journal of Educational Computing Research*, 42, 173-196.
- Kukich, K. (2000, September/October). Beyond Automated Essay Scoring. In M. A. Hearst (Ed.), *The debate on automated essay grading IEEE Intelligent Systems*, 27-31.
- Lai, Y. (2010). Which do students prefer to evaluate their essays: Peers or computer program. *British Journal of Educational Technology*, 41(3), 432-454.
- Landauer, T.K., Laham, D., & Foltz, P.W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In M.D. Shermis & J.C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 87-112). Mahwah, New Jersey: Erlbaum.
- MacArthur, C. A. (2000). New tools for writing: Assistive technology for students with writing difficulties. *Topics in Language Disorders*, 20, 85-100.
- MacArthur, C.A. (2007). Best practices in teaching evaluation and revision. In S. Graham, C.A. MacArthur, and J. Fitzgerald (Eds.), *Best Practices in Writing Instruction* (pp. 141-163). New York: Guilford Press.
- MacArthur, C. A. (2015). Instruction in evaluation and revision. In MacArthur, C. A., Graham, S., & Fitzgerald, J. (Eds.), *Handbook of writing research, 2nd Ed.* (272-287). New York: Guilford.

- MacArthur, C. A. (2012). Evaluation and revision processes in writing. In Berninger, V. W. (Ed.), *Past, Present, and Future Contributions of Cognitive Writing Research to Cognitive Psychology*, (pp. 461-483). London: Psychology Press.
- MacArthur, C.A., Graham, S., Schwartz, S. (1991). Knowledge of revision and revising behaviors among students with learning disabilities. *Learning Disability Quarterly*, 14, 61-73.
- Page, E. (2003). Project Essay Grade: PEG. In M.D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 43-54). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Prior, P. (2003). Tracing process: How texts come into being. In C. Bazerman & P. Prior (Eds.), *What Writing Does and How it Does it: An introduction to analyzing texts and textual practices*. (pp. 167-200). Mahwah, NJ: Erlbaum.
- Scharber, C., Dexter, S., Riedel, E. (2008). Students' experiences with an automated essay scorer. *The Journal of Technology, Learning, and Assessment*, 7(1), 1-45.
- Schultz, M.T. (2013). The IntelliMetric Automated Essay Scoring Engine—A review and an application to Chinese essay scoring. In M.D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 89-98). New York: Routledge.
- Shermis, M.D. & Burstein, J. (2003). Introduction. In M.D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. xiii-xvi). Mahwah, NJ: Lawrence Erlbaum Associates.
- Shermis, M.D., Burstein, J., & Leacock, C. (2006). Applications of computers in assessment and analysis of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 403-416). New York: Guilford.
- Shermis, M.D., Burstein, J., Bursky. (2013). Introduction to automated essay evaluation. In M.D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 1-15). New York: Routledge.
- Shermis, M.D., & Burstein, J. (Eds.). (2013). *Handbook of Automated Essay Evaluation: Current applications and new directions*. New York: Routledge.
- Shermis, M. D., Wilson Garvan, C., & Diao, Y. (2008). *The impact of automated essay scoring on writing outcomes*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.
- Strauss, A. & Corbin, J.C. (1998). *Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory* (2nd Ed.). London: Sage.
- Troia, G. A. (2006). Writing instruction for students with learning disabilities. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 324-336). New York, NY: Guilford.
- U. S. Department of Education, Office of the Deputy Secretary, Implementation and Support Unit. (2012a). *Race to the Top assessment: Partnership for Assessment of Readiness for College and Careers year one report*. Washington, DC: Author. Downloaded from: <http://www2.ed.gov/programs/racetothetop-assessment/performance.html>
- U. S. Department of Education, Office of the Deputy Secretary, Implementation and Support Unit. (2012b). *Race to the Top assessment: Smarter Balanced Assessment Consortium year one report*. Washington, DC: Author. Downloaded from: <http://www2.ed.gov/programs/racetothetop-assessment/performance.html>
- Vantage Learning (2008). MY Access! School Edition [Software]. Available from <http://www.vantagelearning.com/products/my-access-school-edition/>
- Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, 22(3), 333-362.
- Whithaus, C. (2013). Foreword. In M.D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. vii-ix). New York: Routledge.
- Wilson, J., Olinghouse, N. Andrada, G.N. (2014). Does automated feedback improve writing? *Learning Disabilities: A Contemporary Journal*. 12(1), 93-118.

Appendix A: Detailed Description of Data Sources

Think-alouds. Two think alouds were collected from each student. Students worked individually with the primary researcher in a quiet place while revising and thinking aloud (e.g., computer lab, library). Before the first think aloud with students, the researcher explained the meaning of thinking aloud and modeled thinking aloud as she solved a tangram puzzle. Then students practiced thinking aloud using a researcher-constructed task in which students read a brief persuasive letter and thought aloud about how they would help the author improve the writing. Next, the students were told to log on to the AEE program, submit their draft (previously written using the word processor in MY Access!), view their scores, and use the AEE feedback to make revisions. Students were asked to think aloud as they completed these steps. The think-aloud data was recorded using CamStudio. The researcher reminded students to think aloud if they fell silent for longer than 5 seconds. See Appendix B for think aloud procedures.

Semi-Structured Interviews. Immediately following each think aloud, the researcher asked each student a series of questions to confirm what the student did during the think aloud and to gain a broader perspective into their thoughts and perceptions regarding the use of the technology to revise. The interviews were audio recorded. See Appendix C for interview questions.

Student Writing. Before and after drafts of student writing completed during each think aloud were also collected. This resulted in a total of 24 written drafts (i.e., 4 drafts from each student). The revisions made during the think alouds were identified, coded and analyzed.

Appendix B: Think Aloud Procedures

Step 1: Set up computer with microphone and mouse. Try one brief Camstudio video. Play back to make sure it is working properly.

Step 2: Say to student, “As you know, we are interested in how students use MY Access!. I would like you to submit your essay, look at the revision plan or MY Tutor feedback and think about it. You should choose one or a few suggestions from the revision plan or MY Tutor and use these to make revisions to your paper. While you are doing this, I want you to “think aloud”, that is, just tell me everything that you are thinking as you work. For example, tell me what you are doing and how you are feeling; tell me what you are reading; tell me what you think about the scores and the MY Tutor feedback and what you might do with it; tell me any thought that comes into your head: questions that come to mind, plans you might be making, expectations, reactions, feelings, or memories. Remember, there are no wrong or right answers. We just want to know how you use it. It’s even OK if you think about things that aren’t about MY Access!. I’ll give you a little demonstration, using a puzzle called a tangram.

Step 3: Give brief demo of a think aloud using tangram.

Step 4: Ask student to try a practice think aloud with you. Open up the word document “Summer School Task.” Say, “I would like you to practice thinking aloud. This is a letter that a student like you wrote to his principal. He is trying to persuade the principal at his school that they should not have summer school. Pretend that this student asked you to give him some advice to make his writing better. You are going to read his letter and provide advice or feedback to him to help him make his letter better. While you are doing this, I want you to “think aloud”, that is, just tell me everything that you are thinking as you work. For example, tell me what you are doing and how you are feeling; tell me what you are reading; tell me any thought that comes into your head: questions that come to mind, advice you would give, plans you might be making, expectations, reactions, feelings, or memories. Remember, there are no wrong or right answers.

Step 5: Have student log on to MY Access! program.

Step 6: Repeat directions aloud to student, “I would like you to submit your essay, look at the revision plan or MY Tutor feedback and think about it. You should choose one or a few suggestions from the revision plan or MY Tutor and use these to make revisions to your paper. While you are doing this, I want you to “think aloud”, that is, just tell me everything that you are thinking as you work. For example, tell me what you are doing and how you are feeling; tell me what you are reading; tell me what you think about the scores and the MY Tutor feedback and what you might do with it; tell me any

thought that comes into your head: questions that come to mind, plans you might be making, expectations, reactions, feelings, or memories. Remember, there are no wrong or right answers. We just want to know how you use it. It's even OK if you think about things that aren't about MY Access!"

Step 7: Begin Recording. During the TA, give non-directive prompts if the student does things without commenting (e.g., reads the feedback) or falls silent for more than a minute. "Please remember to tell me what you are thinking." "Can you tell me why you are doing that?; What are you reading now? What are you doing? What are you thinking?" If student does not know where to go, direct him. If student begins reading all of the feedback, remind student that he should choose only one or a few suggestions to focus on for this revision.

Step 8: Ask questions after the session. This will be a structured interview with a few questions that will be asked of everyone and flexibility to ask follow-up questions to individual students.

Appendix C: Semi-Structured Interview

General

- 1) What do you think makes a(n) (narrative, persuasive, informative, etc.) essay good? What qualities must it have? How did you learn that?
- 2) When you are writing to a MY Access! prompt, how do you know it's time to/you are ready submit?

Understanding the feedback:

- 1) Let' look at your scores. Can you explain what your scores mean? For example, why do you think you have 6 scores?)
- 2) Do you think the scores are about right for your paper? Do you think your teacher will agree with these scores?
- 3) Is there one score that you pay attention to more than another? Are all scores are equally important? Why?
- 4) Do you usually get the same scores on your essays?
- 5) What did the MY Tutor feedback tell you? What does that mean?
- 6) What suggestions did MY Access! give you and do you think they were helpful?

Revising:

- 1) Did you use the feedback to help you make revisions today or did you use it to think about revisions you will make?
- 2) Point out a particular change: Why did you make this change? Did MA help you to improve your paper?
- 3) Do/did you think you will get a higher score after making these revisions?
- 4) How can someone get a 6 on MY Access!?

Classroom and teacher:

- 1) What kind of writing advice does your teacher give you to help you? What kind of writing advice do the other students in your class give you?
- 2) Is the feedback you get from MY Access! the same sort of feedback that you get from the teacher or other students in your class?

AEE Technology

- 1) How do you think MY Access! knows what score to give you? How does it work?
- 2) Do you like using MY Access! - Why? Why not?

Appendix D: Definitions of Revision Codes

Code	Definition
Surface Revisions	Editing changes involving words, letters, punctuation, and grammar
Type	
<i>Spelling</i>	Changing the spelling of a word
<i>Punctuation</i>	Changing the punctuation (e.g., period, comma)
<i>Capitalization</i>	Changing letter case
<i>Grammar</i>	A change involving a grammatical rule (i.e., subject-verb agreement, possessive)
Quality	
<i>No Success</i>	No change or worse
<i>Success</i>	Improvement
Reason for revision	The reason the change was made as described in student's think aloud protocol: add information or details, clarify, engage audience, focusing, vary word choice, make complete sentence, transition.
Non-Surface Revisions	Substantial changes involving words, phrases, or sentences
Unit size	
<i>T-unit</i>	Change involving one main clause and its dependent clauses
<i>Word</i>	Change involving a word
<i>Phrase</i>	Change involving several words, but not a complete sentence.
Operation	The action required to make the change
<i>Add</i>	Inserting a word, several words or a complete T-unit(s) that is an addition to the text
<i>Delete</i>	Deleting a word, several words or a complete T-unit without replacing it
<i>Rewrite</i>	Inserting a word, several words, or T-unit(s) in place of an existing word, several words, or T-unit(s)
Quality	The extent to which a change improves the quality of the text
<i>No success</i>	The change has not effect
<i>Minor success</i>	The change clarifies an existing idea or gives more information about an existing idea
<i>Major success</i>	The change adds a new idea to the text or sheds new light on an existing idea
Reason for revision	The purpose for making the change as detailed in a student's think aloud protocol
Source prompting revision	The entity prompting the change as detailed in a student's think aloud protocol