Using PEGWriting® to Support the Writing Motivation and Writing Quality of Eighth-Grade Students: A Quasi-Experimental Study

Joshua Wilson
University of Delaware
213E Willard Hall
Newark, DE 19716
joshwils@udel.edu

Trish Martin
Measurement Incorporated
423 Morris Street
Durham, NC 27701
tmartin@measinc.com

Abstract

A quasi-experimental study compared the effects of feedback condition on eighth-grade students' writing motivation and writing achievement. Four classes of eighth-graders were assigned to a combined feedback condition in which they received feedback on their writing from their teacher and from an automated essay evaluation (AEE) system called PEGWriting®. Four other eighth-grade classes were assigned to a teacher feedback condition, in which they solely received feedback from their teacher via GoogleDocs. Results indicated that students in the combined PEGWriting+Teacher Feedback condition received feedback more quickly and indicated that they were more likely to solve problems in their writing. Equal writing quality was achieved between feedback groups even though teachers in the PEGWriting condition spent less time providing feedback to students than in the GoogleDocs condition. Results suggest that PEGWriting enabled teachers to offload certain aspects of the feedback process and promoted greater independence and persistence for students.

1. Introduction

In the 21st century, possessing strong writing skills is essential for success in K-12 education, college acceptance and completion, and stable gainful employment (National Commission on Writing, 2004, 2005). Yet, more than two-thirds of students in grades four, eight, and twelve fail to achieve grade-level proficiency in writing, as indicated by recent performance on the National Assessment of Educational Progress (NCES, 2012; Salahu-Din, Persky, & Miller, 2008). Without sufficient writing skills, students are at-risk of performing worse in school, suffering lower grades, and experiencing school dropout (Graham & Perin, 2007).

One effective method for improving students' writing skills is providing instructional feedback (Graham, McKeown, Kiuhara, & Harris, 2012; Graham & Perin, 2007). Struggling writers, in particular, need targeted instructional feedback because they tend to produce shorter, less-developed, and more error-filled texts than their peers (Troia, 2006). However, instructional feedback is often difficult and time-consuming for teachers to provide. Indeed, educators in the primary and secondary grades report that the time-costs of evaluating writing are so prohibitive that they rarely assign more than one or two paragraphs of writing (Cutler & Graham, 2008; Kiuhara, Graham, & Hawken, 2009). Consequently, educators are increasingly relying on automated essay evaluation (AEE) systems (Warschauer & Grimes, 2008) to provide students with immediate feedback in the form of essay ratings and individualized suggestions for improving an essay—i.e., automated feedback.

Previous research on AEE indicates that, in isolation of teacher feedback, automated feedback appears to support modest improvements in students' writing quality. Findings from studies of ETS's Criterion® (Kellogg, Whiteford, & Quinlan; Shermis, Wilson Garvan, & Diao, 2008), Pearson's Summary Street (Franzke, Kintsch, Caccamise, Johnson, & Dooley, 2005; Wade-Stein & Kintsch, 2004), and Measurement Incorporated's PEGWriting® system (Wilson & Andrada, in press; Wilson, Olinghouse, & Andrada, 2014), indicate that automated feedback assists students in improving the overall quality of their essays while concomitantly reducing the frequency of their mechanical errors.

Less research has explored the effects of AEE on writing motivation. However, in two studies, Warschauer and Grimes (2008; 2010), found that

179

teachers and students who had used ETS' Criterion or Pearson's My Access programs, agreed that AEE had positive effects on student motivation. Teachers also reported that the AEE systems saved them time on grading, and to be more selective about the feedback they gave.

1.1 Study purpose

The purpose of the present study was to extend previous research in the following ways. First, previous studies of AEE have focused on the use of automated feedback in isolation of teacher feedback, despite the intended use of such systems for complementing, not replacing, teacher feedback (Kellogg et al., 2010). To date, no research has evaluated the effects of a combined AEE-and-teacher-feedback condition against a teacher-feedback-only condition. Furthermore, studies have employed a weak control condition, typically a no-feedback condition, to test the effects of AEE on writing quality.

Furthermore, additional research is needed regarding the possible effects of AEE on writing motivation. Theoretical models of writing (e.g., Hayes, 2006; 2012), and empirical research (e.g., Graham, Berninger, & Fan, 2007) underscore the importance of a student's motivation and dispositions towards writing for promoting writing achievement. As AEE systems become more widely-used, it is important for stakeholders to know the degree, and limitations, of their effect on these affective dimensions of writing ability.

Therefore, the present study compared a combined teacher-plus-AEE feedback condition to a teacher-feedback-only condition with regards to their effect on eighth-grade students' writing motivation and writing quality. The combined feedback condition utilized an AEE system called PEGWriting. The teacher-feedback-only condition utilized the comments function of GoogleDocs to provide students with feedback. We hypothesized that students in the combined feedback condition would report greater motivation due to PEGWriting's capacity to provide immediate feedback in the form of essay ratings and individualized suggestions for feedback. With respect to writing quality, it was difficult to generate a priori hypotheses given the aforementioned

limitations of previous research. Exploratory analyses considered whether students in the combined feedback condition outperformed their peers on measures of writing quality, or whether quality was commensurate across groups.

2. Methods

2.1 Setting and Participants

This study was conducted in a middle school in an urban school district in the mid-Atlantic region of the United States. The district serves approx. 10,000 students in 10 elementary schools, three middle schools, and one high school. In this district, 43% of students are African-American, 20% are Hispanic/Latino, and 33% White. Approximately 9% of students are English Language Learners, and 50% of students come from low income families.

Two eighth-grade English Language Arts (ELA) teachers agreed to participate in this study. The teachers were experienced, having taught for a total of 12 and 19 years, respectively. One teacher had earned a Master's degree and the other was in the process of earning it (Bachelor's +21 credits). Each teacher taught a total of four class periods of ELA per day.

Given that the school did not use academic tracking and each class exhibited a range of reading and writing ability, teachers were asked to randomly select two classes from each of their schedules to assign to the combined automated-and-teacher-feedback condition (hereafter referred to as PEG+Teacher), and two classes to assign to a teacher-feedback-only condition (hereafter referred to as GoogleDocs). Thus, teachers instructed classes assigned to both feedback condition.

A total of 74 students were assigned to the PEG+Teacher condition and 77 students to the GoogleDocs condition. Though classes were randomly assigned to feedback conditions, the study sampled intact classes, resulting in a quasi-experimental design. Table 1 reports demographics for each sample. Chi-Square and *t*-tests confirmed that the groups were equal with respect to all variables. In addition, all students received free-lunch. No students received special education services.

	PEG + Teacher	GoogleDocs
Gender (*n*)		
Male	41	38
Female	33	39
Race (*n*)		
Hispanic/Latino	20	20
African American	31	24
White	22	30
Asian	1	1
Unreported	0	2
ELL (*n*)	2	0
Age (months)		
M	169.03	169.51
SD	5.90	4.90

Table 1: Demographics of Study Participants

2.2 Description of PEGWriting

PEGWriting is a web-based formative writing assessment program developed by Measurement Incorporated (MI). It is designed to provide students and teachers with an efficient and reliable method of scoring student writing in order to promote students' writing skills.

PEGWriting is built around an automated essay scoring engine called PEG, or Project Essay Grade. PEG was developed by Ellis Batten Page (Page, 1966; 1994; 2003) and acquired by MI in 2002. PEG uses a combination of techniques such as natural language processing, syntactic analysis, and semantic analysis to measure more than 500 variables that are combined in a regression-based algorithm that predicts human holistic and analytic essay ratings. A number of empirical studies have established the reliability and criterion validity of PEG's essay ratings (Kieth, 2003; Shermis, 2014; Shermis, Koch, Page, Keith, & Harrington, 2002).

Students and teachers access PEGWriting by visiting www.pegwriting.com and inputting their individual username and passwords. Teachers can assign system-created prompts in narrative, argumentative, or informative genres. They can also create and embed their own prompts, which can use words, documents, images, videos, or even music as stimuli.

Once a prompt is assigned, students can select from several embedded graphic organizers to support their brainstorming and prewriting activities. After prewriting, students have up to 60 minutes to complete and submit their drafts for evaluation by PEG. Once submitted, students immediately receive essay ratings for six traits of writing ability: idea development, organization, style, sentence structure, word choice, and conventions. Each of these traits is scored on a 1-5 scale and combined to form an Overall Score ranging from 6-30. In addition, students receive feedback on grammar and spelling, as well as trait-specific feedback that encourages students to review and evaluate their text with regard to the features of that specific trait. Students also receive customized links to PEGWriting's skill-building mini-lessons. These lessons are multimedia interactive lessons on specific writing skills such as elaboration, organization, or sentence variety.

Once students receive their feedback, they may revise and resubmit their essays up to a total of 99 times—the default limit is 30—and receive new essay ratings, error corrections, and trait-specific feedback. Teachers are also able to provide students with feedback by embedding comments within the students' essays or through summary comments located in a text box following the PEG-generated trait-specific feedback. Students may also leave comments for their teacher using a similar function.

2.3 Study Procedures

After classes were assigned to feedback conditions, all students completed a pretest writing motivation survey (Piazza & Siebert, 2008; see Section 2.4). Then, teachers began instruction in their district-assigned curriculum module on memoir writing. Teachers introduced the key features of memoir writing to all their classes. During this initial instructional phase, students in the PEG+Teacher condition were given an opportunity to learn how to use PEGWriting. Earlier in the school year, the first author trained the two teachers on the use PEGWriting during three 30 minute training sessions. Then, teachers subsequently trained their students how to use the program in one 45 minute class period following completion of the pretest writing motivation survey.

Teachers then assigned their district-created writing prompt for the memoir unit, which read:

> We have all had interesting life experiences.
> Some are good, funny, or exciting, while

181

others are bad, sad, or devastating. Choose one experience from your life and tell the story. Once you have chosen your topic, you may choose to turn it into a scary story, drama, elaborate fiction, science fiction, comedy, or just tell it how it is. Be sure to organize your story and elaborate on your details. Your audience wasn't there so you need to tell them every little detail.

Students then proceeded to brainstorm, organize their ideas, and draft their memoirs using the technology available to them. Students in the PEG+Teacher condition used the built-in graphic organizers to plan out their memoirs. Students in the GoogleDocs condition used teacher-provided graphic organizers. Subsequent class periods were devoted to drafting, revising, and editing the memoirs. During this time, teachers delivered mini-lessons on features of memoir writing such as "Show, not tell," "Using dialogue in memoirs," and "Using transitions." Both teachers kept a log of their instructional activities, documenting that they delivered the same instruction as each other and to each of the classes they taught.

Teachers were instructed to review and provide feedback on their students' writing a minimum of one, and a maximum of two times, across both conditions. Teachers were allowed to provide feedback as they normally would, commenting on those aspects of students' text which they deemed necessary. They gave feedback to students in the GoogleDocs condition by (a) directly editing students' texts, and (b) providing comments similar to the comment feature in Microsoft Word. Since students in the PEG+Teacher feedback condition were already receiving feedback from PEG, teachers could supplement the feedback with additional comments as they deemed necessary. Feedback was delivered in the form of embedded comments (similar to the GoogleDocs condition) and in the form of summary comments. Students in this condition were allowed to receive as much feedback from PEG as they wished by revising and resubmitting their memoir to PEG for evaluation. But, the amount of teacher feedback was held constant across conditions.

At the conclusion of the instructional period (approx. three weeks), students submitted the final drafts of their memoir. Then, students were administered a post-test writing motivation survey that mirrored the initial survey with additional items that specifically asked about their perceptions of the feedback they received. Teachers also completed a brief survey regarding their experiences providing feedback via PEGWriting and GoogleDocs.

2.4 Study Measures

Writing Motivation was assessed using the Writing Disposition Scale (WDS; Piazza & Siebert, 2008), which consisted of 11 Likert-scale items that are combined to form three subscales measuring the constructs of confidence, persistence, and passion. Cronbach's Alpha was reported as .89 for the entire instrument, and .81, .75, and .91, respectively for the three subscales (Piazza & Siebert, 2008). The WDS was administered at pretest and at posttest. The posttest administration of the WDS also include additional researcher-developed items asking students to share their perceptions of the feedback they received. These items included Likert-scale ratings followed by an open-ended response option.

Writing quality was assessed using the PEG Overall Score, PEG trait scores, and teacher grades. Details on the PEG Overall Score and the PEG trait scores are found in Section 2.2. Teacher grades were generated by using a primary trait narrative rubric developed by the local school district. The rubric evaluated ten traits of personal narrative writing, each on a 0-10 scale. Final grades were assigned by totaling students' scores on each of the ten traits (range: 0-100). Traits assessed included: the presence of a compelling introduction; logical organization; establishment of a setting, narrator, and point of view; effective conclusion which reflects on the life event; sufficient details and description; effective use of figurative language and dialogue; presence of accurate sentence structure; strong and vivid word choice; and absence of errors of spelling, punctuation, and usage.

2.5 Data Analysis

Non-parametric analyses were used to estimate differences between feedback conditions on individual items of the Writing Dispositions Scale (WDS). A series of one-way analysis of variance (ANOVA) were used to estimate differences between groups on the Confidence, Persistence, and

	PEG+Teacher					GoogleDocs				
Item	SA	A	N	D	SD	SA	A	N	D	SD
1. My written work is among the best in the class.	8	17	34	13	2	9	13	39	10	5
2. Writing is fun for me.	4	25	29	8	8	10	23	15	18	10
3. I take time to try different possibilities in my writing.	3	33	23	13	2	7	35	20	10	4
4. I would like to write more in school.	2	11	25	22	14	5	18	18	17	18
5. I am NOT a good writer.	3	12	23	20	16	6	8	28	25	9
6. Writing is my favorite subject in school.	3	6	24	31	10	3	11	22	22	18
7. I am will to spend time on long papers.	3	21	21	14	15	8	23	13	19	13
8. If I have choices during free time, I usually select writing.	0	4	13	27	30	1	8	8	27	32
9. I always look forward to writing class.	3	10	27	21	13	1	10	24	18	23
10. I take time to solve problems in my writing.	11	34	17	8	4	5	34	20	13	4
11. Writing is easy for me.	10	24	31	2	7	17	22	27	5	5

Table 2: Frequencies of Student Responses to the Pretest Writing Disposition Scale (WDS) by Feedback Condition
SA = Strongly Agree; A = Agree; N = Neutral; D = Disagree; SD = Strongly Disagree

Passion subscales. Confidence was formed as the average of items 1, 5 (reverse coded), and 11. Reverse coding was achieved by translating self-reports of strongly agree to strongly disagree, agree to disagree, and vice versa. Neutral responses remained the same. Persistence was formed as the average of items 3, 4, 7, and 10. Passion was formed as the average of items 2, 6, 8, and 9. Finally, a series of one-way ANOVAs was used to compare conditions with respect to the writing quality measures. Full data was available for all students on the PEG Overall and Trait scores. Data on teacher grades was only available for 62 students in each group at the time of this reporting. Data coded from open-ended response items from teachers and students was used to contextualize the results. Missing data for posttest measures of motivation and writing quality resulted in listwise deletion of cases from analyses

3. Results

3.1 Pretest Analyses of Writing Motivation

Data from the pretest administration of the WDS is presented in Table 2 (above). Non-parametric analyses performed on the individual survey items revealed that the null hypothesis of equal distributions across feedback conditions was retained in all cases. Thus, it is possible to assume that students' writing motivation did not differ as a function of their feedback condition. Means and standard deviations for the subscales of Confidence, Persistence, and Passion are presented in Table 3. T-tests indicated no-statistically significant differences in subscale scores by feedback condition. Hence, at pretest, groups were equivalent with respect to their writing motivation and writing dispositions.

	PEG+Teacher		GoogleDocs	
Subscale	*M*	*SD*	*M*	*SD*
Confidence	2.75	.91	2.58	.78
Persistence	3.05	.85	2.86	.69
Passion	3.64	.88	3.42	.76

Table 3: Descriptives for WDS Subscales at Pretest

3.2 Posttest Analyses of Writing Motivation

Non-parametric analyses were performed on the individual posttest survey items, examining

statistically significant differences in the distribution of responses across conditions. All contrasts were non-statistically significant, except for item 10—"I take time to solve problems in my writing." The mean ranks of the PEG+Teacher and GoogleDocs conditions were 62.52 and 75.53, respectively: $U = 1921.50$, $Z = -2.03$, $p = .04$. Examination of the individual frequency data for this item (see Table 4) shows that 66% of students in the PEG+Teacher feedback condition agreed or strongly agreed with this statement, as compared to 50% of students in the GoogleDocs condition.

	SA	A	N	D	SD
PEG+Teacher	11	31	17	4	1
GoogleDocs	7	30	27	7	3

Table 4: Posttest Frequencies to WDS Item 10

When comparing pre-/posttest responses to item 10 (see Figure 1), the percentage of students in the PEG+Teacher condition who agreed or strongly agreed that they take time to solve problems in their writing increased by 5%, whereas those in the GoogleDocs condition stayed the same. One-way ANOVAs comparing subscale scores across condition were not statistically significant.
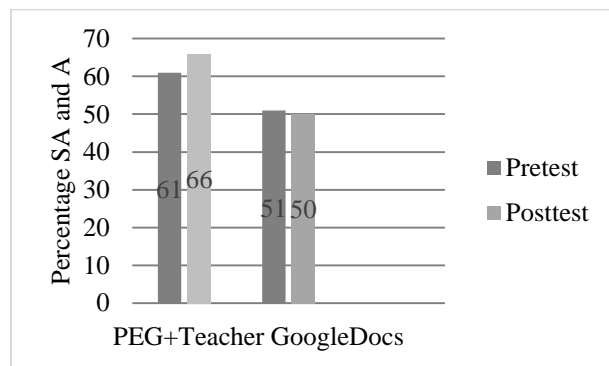

Figure 1: Pretest/Posttest Comparison of SA/A Responses to WDS Item 10

To further investigate this finding we compared the average number of essay drafts completed by students in each condition using a one-way ANOVA. Results indicated that students in the PEG+Teacher condition completed a higher average number of essay drafts ($M = 11.28$, $SD = 6.81$) than students in the GoogleDocs condition ($M = 7.18$, $SD = 2.29$): $F (1, 138) = 22.287$, $p < .001$.

Thus, students' self-report information and behavior appears to be consistent in this regard.

In addition to the 11 items on the WDS scale, seven other Likert-scale items were administered at posttest assessing students' perceptions of the feedback they received. Non-parametric analyses indicated a statistically significant difference between feedback conditions on item 18—"I received feedback quickly"—favoring the PEG+Teacher feedback condition (Mean rank = 56.34) as compared to the GoogleDocs condition (Mean rank = 79.31): $U = 1526.00$, $Z = -3.57$, $p < .001$. A total of 78% of students in the PEG+Teacher condition agreed or strongly agreed that they received feedback quickly, as compared to 63% of students in the GoogleDocs condition. Examination of the frequency data for the other feedback-specific items (see Table 5 following page) suggests that students in both conditions perceived feedback to be useful for helping them improve their writing. Students exhibited greater variation with regard to their desire to receive more feedback (Item 15). Open-ended response data suggests that feedback can serve to both encourage and discourage student writers. For some, feedback is a supportive and motivating factor.

> I wish that because it'll help me make my writing pieces better. Than that way I know what to do and what mistakes not to do next time. (ID #2324)

Yet for others, feedback serves to highlight a student's deficits and cause discomfort.

> I got tired of so much feedback. (ID #2403)
> Not really because I wouldn't like people to know what I'm writing. (ID #2301)

Still, for some students it is the absence of feedback, not the presence of it, which tells them that they are doing a good job.

> I chose three because yes I would like to receive feedback but if I don't I think I'm doing fine. (ID #2321)

| | PEG+Teacher | | | | | GoogleDocs | | | | |
Item	SA	A	N	D	SD	SA	A	N	D	SD
12. The Feedback I received helped me improve my writing.	27	31	4	2	0	26	42	4	0	1
13. I received the right amount of feedback.	24	27	12	1	0	23	36	11	2	1
14. The feedback I received made sense to me.	23	29	7	4	0	25	39	8	1	0
15. I wish I had more opportunities to receive feedback.	14	15	16	14	4	17	19	20	11	6
16. I received feedback about a variety of writing skills.	15	32	9	8	0	14	27	20	9	3
17. Receiving feedback on my essay score helped me improve my writing.	33	25	8	4	2	34	27	7	3	2
18. I received feedback quickly.	30	20	12	2	0	12	33	17	10	0

Table 5: Frequencies by Condition of Student Responses to the Posttest Survey Items Regarding Feedback

3.3 Posttest Analyses of Writing Quality

A series of one-way ANOVAs examined the effects of feedback condition on the PEG Overall Score and PEG trait scores. The null hypothesis of equal means was retained in all cases. However, the one-way ANOVA comparing groups on the "Conventions" trait approached statistical significance, showing a small effect size favoring the PEGWriting group: $F(1, 138) = 3.33$, $p = .07$, $D = .31$. There was also a small effect size favoring the PEGWriting condition on the Sentence Structure trait: $D = .18$. The one-way ANOVA of Teacher Grades was not statistically significant, but a small effect size favored the PEGWriting group: $D = .19$.

3.4 Teacher Survey Data

Results of surveys administered to teachers at the conclusion of the study indicated that teachers varied their feedback across conditions. Teachers were asked to rank the following skills in order of the frequency with which they were commented on in students' writing: spelling, punctuation, capitalization, organization, idea development and elaboration, and word choice. For the GoogleDocs condition, teachers agreed that they most frequently provided feedback on low-level writing skills, such as spelling, punctuation, capitalization, and grammar. For the PEG+Teacher condition, teachers agreed that they most frequently provided feedback on high-level writing skills: idea development and

Writing Skills	GoogleDocs	PEG+Writing
Capitalization	✓	
Grammar	✓	
Idea Development & Elaboration		✓
Organization		✓
Punctuation	✓	
Spelling	✓	
Word Choice		✓

Figure 2. Writing Skill Feedback by Condition

elaboration, organization, and word choice. Indeed, one teacher said she did not need to give any feedback on capitalization or grammar. When asked to decide which of the two systems—PEGWriting or GoogleDocs—enabled them to devote more energy to commenting on content, both teachers selected PEGWriting.

Teachers further agreed that they needed to give less feedback to students who had been using PEGWriting. Consequently, when asked to estimate the amount of time spent providing feedback to students in each condition, teachers agreed that providing feedback in the GoogleDocs condition took twice as long as doing so in the PEG+Teacher condition. For this reason, both teachers agreed that PEGWriting was more efficient for providing feedback than GoogleDocs.

When asked to select which system was easier for teachers and students to use, teachers agreed that GoogleDocs was easier for teachers, but PEGWriting and GoogleDocs were equally easy for students to use. However, both teachers agreed that PEGWriting was more motivating for students and that it promoted greater student independence.

4. Discussion

This study was the first of its kind to compare the effects of a combined automated feedback and teacher feedback condition and a teacher-feedback-only condition (GoogleDocs) on writing motivation and writing quality. Students in the combined feedback condition composed memoirs with the aid of feedback from an AEE system called PEGWriting® and their teacher (provided within the environment of PEGWriting). Students in the teacher-feedback-only condition composed their texts using GoogleDocs which enabled their teacher to edit their text and embed comments.

Based on prior research (Grimes & Warschauer, 2010; Warschauer & Grimes, 2008), we hypothesized that students would report greater writing motivation in the PEG+Teacher feedback condition. However, we were unable to generate an a priori hypothesis regarding the effects of feedback condition on writing quality since prior research has not contrasted feedback conditions in the manner investigated in the current study.

With respect to writing motivation, our hypothesis was partially confirmed. Students in the PEG+Teacher feedback condition reported stronger agreement with Item 10 of the WDS—"I take time to solve problems in my writing"—than did students in the GoogleDocs condition. This self-report data was confirmed with a statistically significant difference, favoring the PEG+Teacher feedback condition, in the number of drafts students completed. However, effects on broader constructs of writing motivation—confidence, persistence, and passion—were not found. This may have been due, in part, to the duration of the study. The study spanned just over three weeks; hence, it is likely that additional exposure and engagement with PEGWriting is needed to register effects on these broader constructs.

Nevertheless, it is encouraging that students reported greater agreement with Item 10. Revision is a challenging and cognitively-demanding task

(Flower & Hayes, 1980; Hayes, 2012), requiring students to re-read, evaluate, diagnose, and select the appropriate action to repair the problem. Many struggling writers lack the motivation to engage in this process, and consequently make few revisions to their text (MacArthur, Graham, & Schwartz, 1991; Troia, 2006). Perhaps, the use of an AEE system, such as PEGWriting, provides sufficient motivation for students to persist in the face of the substantial cognitive demands of revision. Future research should explore the use of AEE systems over extended timeframes. It may be possible that these initial gains in persistence are leveraged to increase writing motivation more broadly.

With respect to writing quality, results showed no statistically significant differences between conditions for the PEG Overall Score or PEG trait scores. While on this surface this may appear to indicate that the feedback provided by PEGWriting yielded no value-added over simply receiving teacher feedback in the form of edits and comments via GoogleDocs, we do not believe this to be the case.

First, though AEE systems are designed and marketed as supporting and complementing teacher feedback, previous research has solely examined the use of automated feedback in isolation from teacher feedback. Furthermore, prior studies have typically employed weak control conditions, such as a no-feedback condition (Kellogg et al., 2010) or a spelling-and-text-length condition (Franzke et al., 2005; Wade-Stein & Kintsch, 2004). While these studies provide important initial evidence of the effects of automated feedback and AEE, their design lacks ecological validity as they do not reflect the intended use of such systems. Lack of statistically significant effects on our measures of writing quality may simply be due to the presence of a stronger counterfactual. Thus, the presence of a stronger control condition in our study should not be confused with absence of value-added.

Second, results from the additional posttest survey items administered to students (see Table 5) and from the survey administered to teachers may point to where the value added by AEE. The provision of immediate feedback in the form of essay ratings, error correction, and trait-specific feedback appears to have enabled students to increase their persistence and independence in solving problems in their writing. Consequently, teachers spent half the amount of time providing

feedback to students as they did to students in the GoogleDocs condition. Moreover, the use of PEGWriting enabled teachers to devote attention to higher-level skills such as idea development and elaboration, organization, and word choice, while offloading feedback on lower-level skills to the AEE system.

Thus, the value-added of PEGWriting appears to be its ability to promote an equivalent level of writing quality as is achieved using a more time consuming and effortful method of providing feedback (i.e., teacher-feedback-only). In other words, by enabling teachers to be more selective and focused in their comments, PEGWriting saved teachers time and effort without sacrificing the quality of students' writing. Forthcoming analyses will determine whether this hypothesis holds true across other measures of writing quality.

4.1 Limitations and Future Research

Study findings and implications must be interpreted in light of the following limitations. First, though teachers randomly assigned classes to feedback conditions, in absence of a pretest measure of writing ability it is not possible to test whether groups were truly equivalent in terms of prior writing ability. Nevertheless, the pretest measure of writing motivation indicated equivalence across conditions with regards to specific writing dispositions and subscales of confidence, persistence, and passion. It is likely, that if one condition exhibited significantly greater writing achievement this would also have been reflected in the disposition ratings (see Graham et al., 2007).

Second, the study examined the effects of feedback on just a single writing assignment: memoir writing. Furthermore, the prompt allowed for substantial student choice, both in terms of the content of their memoir and the form. Students had the freedom to turn their memoir into a comedy, a drama, a science fiction story, or simply recount the events as they happened. It is unclear whether similar results would have been found had a prompt been assigned that was more restrictive in terms of student choice and that placed greater demands on students in terms of background knowledge. Given the literature on prompt and task effects in writing (Baker, Abedi, Linn, & Niemi, 1995; Chen, Niemi, Wang, Wang, & Mirocha, 2007), it is important that future research attempt to replicate results across different writing tasks.

Finally, the sample was drawn from classes taught by two teachers in a single middle school in a school district in the mid-Atlantic region of the United States. Therefore, it is unclear the degree to which study results reflect generalizable or local trends. Nonetheless, study findings on the utility of AEE are consistent with prior research which has used much larger samples (Warschauer & Grimes, 2008). Further, since the study utilized a novel design—comparing a combined AEE and teacher feedback condition to teacher-feedback-only condition—it is logical to initially test the design using smaller samples. Future research should seek to utilize similar feedback conditions in larger samples.

5. Conclusion

The increasing application of AEE in classroom settings necessitates careful understanding of its effects on students' writing motivation and writing quality. Research should continue to illustrate methods of how AEE can complement, not replace, teacher instruction and teacher feedback. The current study provides initial evidence that when such a combination occurs, teachers save time and effort and they provide greater amounts of feedback relating to students' content and ideas. In addition, students receive feedback more quickly, report increases in their persistence to solve problems in their writing. In sum, AEE may afford the opportunity to shift the balance of energy from teachers to students without sacrificing the final quality of students' writing.

Acknowledgements

References

Baker, E. L., Abedi, J., Linn, R. L., & Niemi, D. (1995). Dimensionality and generalizability of domain independent performance assessments. *Journal of Educational Research, 89*(4), 197-205.

Chen, E., Niemie, D., Wang, J., Wang, H., & Mirocha, J. (2007). *Examining the generalizability of direct writing assessment tasks. CSE Technical Report 718.* Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. *Journal of Educational Psychology, 100*, 907-919.

Flower, L. S., & Hayes, J. R. (1980). The dynamics of composing: making plans and juggling constraints. In L.W. Gregg, & E.R. Sternberg (Eds.), *Cognitive processes in writing* (pp. 3- 29). Hillsdale, NJ: Lawrence Erlbaum Associates.

Franzke, M., Kintsch, E., Caccamise, D., Johnson, N., and Dooley, S. (2005). Summary Street®: Computer support for comprehension and writing. *Journal of Educational Computing Research, 33*, 53–80

Graham, S., Berninger, V., & Fan, W. (2007). The structural relationship between writing attitude and writing achievement in first and third grade students. *Contemporary Educational Psychology, 32*(3), 516-536.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*, 879-896.

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

Grimes, D. & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *Journal of Technology, Learning, and Assessment, 8*(6), 1-44. Retrieved December 12, 2014 from http://www.jtla.org.

Hayes, J. R. (2006). New directions in writing theory. In C. MacArthur, S. Graham, J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 28-40). New York: Guilford Press.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*(3), 369-388.

Keith, T. Z. (2003). Validity and automated essay scoring systems. In M. D. Shermis, & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp.147-167). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students learn to write? *Journal of Educational Computing Research, 42*(2), 173-196.

Kiuhara, S. A., Graham, S., & Hawken, L. S. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology, 101*, 136-160.

MacArthur, C. A., Graham, S., & Schwartz, S. (1991). Knowledge of revision and revising behavior among students with learning disabilities. *Learning Disability Quarterly, 14*, 61-73.

McNamara, D.S., Louwerse, M.M., Cai, Z., & Graesser, A. (2013). *Coh-Metrix version 3.0.* Retrieved from http://cohmetrix.com.

National Center for Education Statistics (2012).*The Nation's Report Card: Writing 2011*(NCES 2012–470). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Commission on Writing for America's Families, Schools, and Colleges. (2004). *Writing: A ticket to work…or a ticket out. A survey of business leaders.* Iowa City, IA: The College Board.

National Commission on Writing for America's Families, Schools, and Colleges. (2005). *Writing: A powerful message from state government.* Iowa City, IA: The College Board.

Page, E. B. (1994). Computer grading of student prose, using modern concepts and software. *The Journal of Experimental Education, 62*(2), 127-142.

Page, E. B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan, 48*, 238-243.

Page, E. B. (2003). Project Essay Grade: PEG. In M. D. Shermis, & J. C. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp.43-54). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Piazza, C. L., & Siebert, C. F. (2008). Development and validation of a writing dispositions scale for elementary and middle school students. *The Journal of Educational Research, 101*(5), 275-286.

Salahu-Din, D., Persky, H., and Miller, J. (2008). *The Nation's Report Card: Writing 2007* (NCES 2008–468). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing, 20*, 53-76.

Shermis, M. D., Koch, C. M., Page, E. B., Keith, T. Z., & Harrington, S. (2002). Trait ratings for automated essay grading. *Educational and Psychological Measurement, 62*, 5-18.

Shermis, M. D., Wilson Garvan, C., & Diao, Y. (2008, March). *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.

188

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.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction, 22*, 333-362.

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal, 3*, 22-36.

Wilson, J., & Andrada, G. N. (in press). Using automated feedback to improve writing quality: Opportunities and challenges. In In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of Research on Computational Tools for Real-World Skill Development.* IGI Global.

Wilson, J., Olinghouse N. G., & Andrada, G. N. (2014). Does automated feedback improve writing quality? *Learning Disabilities: A Contemporary Journal, 12*, 93-118.