Implementing Automated Writing Evaluation in Different Instructional Contexts: A Mixed-Methods Study

Corey Palermo[1,2] & Joshua Wilson[3]

[1] North Carolina State University, NC | USA
[2] Measurement Incorporated, Durham NC, USA
[3] University of Delaware, DE | USA

Abstract: There is increasing evidence that automated writing evaluation (AWE) systems support the teaching and learning of writing in meaningful ways. However, a dearth of research has explored ways that AWE may be integrated within different instructional contexts and examined the associated effects on students' writing performance. This paper describes the AWE system MI Write and presents results of a mixed-methods study that investigated the integration and implementation of AWE with writing instruction at the middle-school level, examining AWE integration within both a traditional process approach to writing instruction and with strategy instruction based on the Self-Regulated Strategy Development model. Both instructional contexts were evaluated with respect to fostering growth in students' first-draft writing quality across successive essays as well as students' and teachers' experiences and perceptions of teaching and learning with AWE. Multilevel model analyses indicated that during an eight-week intervention students in both instructional contexts exhibited growth in first-draft writing performance and at comparable rates. Qualitative analyses of interview data revealed that AWE's influence on instruction was similar across contexts; specifically, the introduction of AWE resulted in both instructional contexts taking on characteristics consistent with a framework for deliberate practice.

Keywords: automated writing evaluation; writing instruction; writing assessment

1.   Introduction and Theoretical Basis

Writing is a complex cognitive skill (Berninger & Swanson, 1994; Flower & Hayes, 1980; Hayes, 2012). As such, it requires sustained deliberate practice for the development of writing expertise (Kellogg & Whiteford, 2009). Sustained deliberate practice involves effortful activity to improve performance, intrinsic motivation for task engagement, and frequent practice opportunities (Ericsson, 2006). Practice of this kind is essential to help students develop fluency and automaticity in lower-level writing skills, like handwriting, keyboarding, spelling, and applying the conventions of written language (Berninger & Swanson, 1994; Kellogg, 2008; McCutchen, 1988). Furthermore, practice of this kind is essential to help students become more strategic and exert metacognitive control over the central cognitive processes involved in composing, namely, planning, translating, reviewing, and revising (Bereiter & Scardamalia, 1987; Flower & Hayes, 1980; Graham, 2018; Graham et al., 2019; Harris, Graham, Brindle, & Sandmel, 2009; Hayes, 1996, 2012).

However, practice alone is insufficient. The development of expertise through sustained deliberate practice also requires frequent, appropriate, and effective feedback from one or more feedback agents such as a teacher, peer, self, or computer (Ericsson, 2006; Hattie & Timperley, 2007). Previous research has suggested that students reap the greatest benefit from feedback that is immediate, specific, localized, and detailed, and that addresses both surface-level and content features of writing (Hattie & Timperley, 2007; Nelson & Schunn, 2009; Patchan, Schunn, & Correnti, 2016; Patthey-Chavez, Matsumura, & Valdés, 2004; Shute, 2008).

Unfortunately, with so little time for writing instruction included in most curricula (Brindle, Graham, Harris, & Hebert, 2015; Gilbert & Graham, 2010; Graham, Harris, Fink-Chorzempa, & MacArthur, 2003), students rarely engage in the manner of sustained deliberate practice just described. Furthermore, the provision of high-quality feedback is demanding in terms of teachers' time (Dikli, 2010) and their pedagogical content knowledge (Parr & Timperley, 2010). Indeed, teacher feedback is often ineffective, focusing on low-level writing skills, which have little effect on students' performance (Clare, Valdés, & Patthey-Chavez, 2000; Matsumura, Patthey-Chavez, Valdés, & Garnier, 2002). Consequently, it is critical to identify methods of increasing writing practice and increasing the frequency and effectiveness of teacher feedback in order to improve students' writing outcomes.

One promising method of increasing writing practice and accelerating the practice-feedback loop (Kellogg, Whiteford, & Quinlan, 2010) without increasing the time costs incurred by teachers when evaluating and commenting on students' work is the use of automated writing evaluation (AWE) systems. AWE systems leverage automated feedback capabilities, often along with automated

scoring and various learning-management functions, to support the teaching and learning of writing. Many educators have begun adopting AWE systems (Palermo & Thomson, 2018; Stevenson, 2016; Wilson & Czik, 2016); however, adoption has outpaced research that explores ways that AWE may be effectively implemented and integrated into different forms of teacher-led writing instruction.

Therefore, the present study adopted a mixed-methods design to examine the implementation and integration of AWE with two different approaches to writing instruction at the middle-school level. One instructional context integrated AWE within a traditional process approach to writing instruction (i.e., instruction that incorporated authentic writing opportunities and cycles of planning drafting, and revision), while the other context integrated AWE within strategy instruction (i.e., explicit instruction on cognitive and metacognitive strategies for executing various writing processes like planning, drafting, and revising). We evaluated both instructional contexts with respect to fostering growth in students' first-draft writing quality across successive essays over time. Further, we examined students' and teachers' experiences with and perceptions of the AWE system to understand benefits and limitations of AWE as perceived by users, and how these perceptions related to and explained students' writing performance trajectories in the two instructional contexts.

1.1 Automated Writing Evaluation

AWE systems are technology-based instructional tools developed to support the teaching and learning of writing. A central feature of AWE is the provision of automated feedback intended to guide improvements in writing quality upon revision (see, in this issue, Cotos, Huffman, & Link, 2020; Knight et al., 2020). In doing so, AWE is intended to increase students' access to cycles of writing practice and feedback while reducing teachers' evaluative demands. Commonly, AWE systems complement automated qualitative feedback with automated quantitative feedback in the form of scores or other evaluation metrics (e.g., Mayfield et al., 2018; Roscoe, Allen, Weston, Crossley, & McNamara, 2014; Roscoe & McNamara, 2013). Thus, AWE as a feedback tool often relies on advances in automated essay scoring (AES)—AES refers to automated scoring algorithms that are trained and validated to reliably mimic the scoring of human raters (Shermis & Hamner, 2013).

Prior research has shown that the use of AWE and automated feedback has a number of benefits for supporting the teaching and learning of writing. With respect to teachers and teaching, AWE has been shown to save teachers time by reducing grading, supporting individualized instruction, facilitating classroom management by increasing student autonomy and motivation for writing, assisting with portfolio management, and supporting teachers in providing greater amounts of feedback on higher-level writing skills (Grimes & Warschauer, 2010; Warschauer & Grimes, 2008; Wilson & Czik, 2016; Wilson & Roscoe, 2020).

However, in some instances, the adoption of AWE has not been accompanied by increases in the number of writing opportunities students experienced (Warschauer & Grimes, 2008). At the middle and secondary levels, the pressures of keeping pace with an English Language Arts (ELA) curriculum that de-emphasizes writing may be stronger than the time-saving affordances of AWE (Wilson & Roscoe, 2020). In addition, though automated feedback has been shown to be effective in scaffolding improvements in writing quality across successive revisions to an essay (Wilson & Czik, 2016; Wilson, Olinghouse, & Andrada, 2014), students have, at times, found automated feedback too extensive and overwhelming (Grimes & Warschauer, 2010; Ranalli, 2018), requiring additional teacher support and instruction to interpret.

With respect to students and learning, the use of AWE has been associated with benefits for a number of writing outcomes, including: improvements in time on task, the amount of revising students completed, and the content of K–12 students' writing (Franzke, Kintsch, Caccamise, Johnson, & Dooley, 2005; Graham, Hebert, & Harris, 2015; Grimes & Warschauer, 2010; Morphy & Graham, 2012; Shermis, Garvan, & Diao, 2008; Wade-Stein & Kintsch, 2004). In addition, a smaller number of studies have found automated feedback associated with improvements in writing attitudes (Roscoe, Allen, Johnson, & McNamara, 2018) and writing motivation and self-efficacy (Grimes & Warschauer, 2010; Wilson & Czik, 2016; Wilson & Roscoe, 2020); improvements in writing quality, particularly mechanical aspects of writing across revisions (Kellogg et al., 2010; Morphy & Graham, 2012; Wilson, 2017; Wilson et al., 2014; Wilson & Czik, 2016); and superior performance on state ELA exams (Wilson & Roscoe, 2020).

Though outcomes of AWE usage are generally positive, insufficient research has examined how AWE might be used and integrated with teacher-led writing instruction for optimal effectiveness (c.f., Knight et al., 2020). For instance, prior research by Roscoe and colleagues has productively examined ways of integrating AWE in different student-directed writing-practice formats, finding that AWE practice formats such as writing-process practice, strategy-based practice, and game-based practice yield similar effects on students' improvements in writing quality when revising their writing (see Roscoe et al., 2018; Roscoe, Snow, & McNamara, 2013). However, these studies do not elucidate the ways in which teachers integrate, adapt, and modify their instruction in response to the use of AWE. Similarly, much of the prior research on AWE that has examined teacher-level contrasts has compared an AWE feedback condition to a teacher feedback condition (Stevenson & Phakiti, 2014). While such contrasts may be useful from a research-design perspective, such contrasts lack ecological validity by setting up a false dichotomy between AWE feedback and teacher feedback. Indeed, automated feedback provided by AWE systems is intended to augment rather than replace teacher feedback (Kellogg et al., 2010), allowing teachers to be more

selective in the feedback they provide (Wilson & Czik, 2016). In sum, additional research is needed to inform educators of which affordances of AWE they may capitalize on, what limitations to expect and how to address them, and how AWE can be integrated within different teacher-led instructional contexts, such as process writing instruction or strategy instruction.

Thus, the present study examines not only the efficacy of automated feedback for supporting generalized improvements in students' writing quality across multiple essays over time, but also examines teachers' and students' perceptions of the use of AWE in the context of a traditional process approach to writing instruction and strategy instruction. In so doing, the present study extends the seminal work of Warschauer and Grimes who evaluated district-wide adoption of AWE more than 10 years ago (Grimes & Warschauer, 2010; Warschauer & Grimes, 2008). As with their research, the present study attends to methods of implementing AWE and includes the voices of those whose perceptions influence the adoption, use, and effectiveness of AWE. The present study is unique in further employing longitudinal growth models to measure the rate of students' growth in writing quality when composing multiple essays in different instructional contexts. Indeed, prior research has most often examined growth in writing quality across drafts within an essay or examined growth from pretest to posttest (see Stevenson & Phakiti, 2014). Insufficient research has examined how AWE might be integrated with teacher-led instruction to support generalized improvements in writing quality, that is, transfer to improved performance on writing assignments completed independently without the support of automated feedback (i.e., improvements in the quality of students' first drafts).

2. MI Write

One prominent AWE system is MI Write, formerly known as PEG Writing, which is developed and managed by Measurement Incorporated. MI Write is intended as a formative assessment tool, supporting the teaching and learning of writing. It is a web-based interactive learning environment that supports a number of interactions between teachers, students, and the AWE system. Teachers can utilize MI Write's learning management tools to create customizable prompts, provide students with embedded (i.e., in-text) or summary comments and feedback, and generate customizable reports to monitor class or student progress. Students can leverage MI Write's feedback and scoring capabilities to increase the amount of revising they do, calibrate their performance against automated writing quality scores, use automated feedback to improve their writing quality across successive drafts to an essay, complete differentiated and interactive multimedia skill-building lessons, and give and receive either anonymous or identifiable peer reviews.

To provide students with automated feedback, the MI Write system utilizes the Project Essay Grade (PEG) scoring engine, which statistically analyzes essays, calculates measures reflecting the intrinsic characteristics of writing, and models the decisions of professional raters in producing scores. PEG relies on natural language-processing techniques (e.g., syntactic parsers, semantic analyses) to identify and extract from the training set of essays text features that have known or emergent correlations with human-scored measures of writing quality. Examples of such features include n-grams of characters, words, parts of speech, and phrases; number of mature words or the average hypernym level of the vocabulary; and measures of semantic overlap and other proxies for organization and elaboration. Based on the text features, PEG provides pertinent automated feedback in the form of suggestions for improving the quality of the essay when revising.

With respect to PEG's automated quantitative feedback, MI Write applies the PEG scoring engine for formative purposes, evaluating student writing based on purpose-specific (i.e., informative, argumentative, narrative) but prompt-general scoring models in order to scaffold revision and improve writing performance. MI Write uses scoring models that output automated scores for six traits of writing quality: development of ideas, organization, style, sentence structure, word choice, and conventions. Each trait is scored on a 1–5 scale and summed to form an Overall Score that ranges from 6 to 30. The trait and overall scores are presented to students via a score report that also includes spelling and grammar feedback annotated on the draft, descriptive evaluation and feedback for each trait, and recommended interactive lessons (see Figures 1–4). Teacher reports provide student- and class-level performance (by trait, overall score, and writing purpose), as well as prompt, usage (for essays, drafts, peer reviews, and lessons), and progress data.

Like other AWE systems that utilize automated scoring, PEG's automated scoring system is 100% consistent (i.e., perfectly reliable) and free from common types of human-rater error, such as rater drift or halo effects. Evidence of PEG's reliability was shown recently in the first two phases of the Automated Student Assessment Prize (ASAP) competition sponsored by the Hewlett Foundation. There, PEG achieved the highest level of agreement with human scores of all competitors, and was found to be more reliable overall than two professional raters (Morgan, Shermis, Van Deventer, & Vander Ark, 2013; Shermis & Hamner, 2013).

In addition to the findings by Wilson and colleagues (Wilson, 2017; Wilson & Czik, 2016; Wilson et al., 2014; Wilson & Roscoe, 2020), a recent study of MI Write bears further description. Palermo and Thomson (2018) examined effects of teacher implementation of the AWE system NC Write (a state-specific version of MI Write) as part of a program of writing instruction on middle school students'
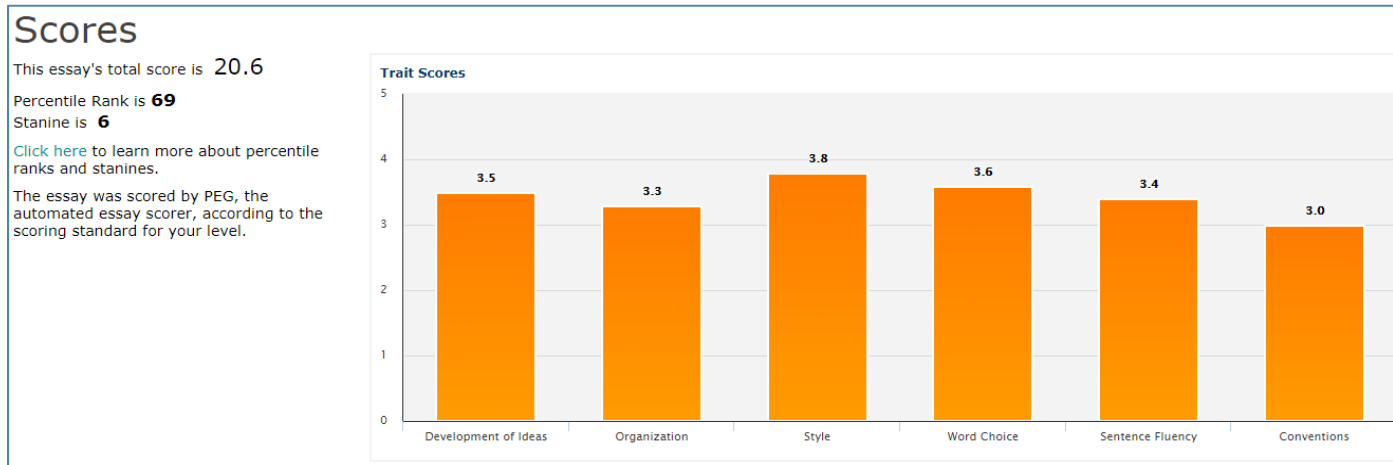
Scores

This essay's total score is **20.6**

Percentile Rank is **69**
Stanine is **6**

Click here to learn more about percentile ranks and stanines.

The essay was scored by PEG, the automated essay scorer, according to the scoring standard for your level.

Trait Scores

Development of Ideas	3.5
Organization	3.3
Style	3.8
Word Choice	3.6
Sentence Fluency	3.4
Conventions	3.0

Figure 1. Essay total and trait-specific scores.

Middle school and high school students should start school later and let us sleep more. We work harder than the elementary and we stay up later than them [G]too.We keep getting more tests [S]then them every year we also get out of school later than them and they go to school later than we [G]do.We should get more sleep because we run harder and do more in P.E. and think about the people who has to do a sport. Think about them who has to do a sport and work a long shift.

All of us should get enough sleep to how much we need. The little kids always get more than us. We have to get up at 5 or 6 while they get up at 6 or 7. And they get more sleep than us because they go to bed earlier than us. The [G]parents would say "it's your fault because you [G]didn't went to bed early enough". Sometimes it's the parent keeping the kids up watching a movie or a tv series. [S]Their are people who have epilepsy. For example, who barely can sleep they could go to bed at 3:30 in the morning and they have to get up at 5:30 in the morning maybe earlier.

Students with [G]jobs_have to [G]work_like you have to work from 4 to 10 p.m. and they go to bed late. So you get better education if you are awake. You don't learn that much if you are tired. You want kids to stay awake and pay attention. [G][S]If they don't pay attention they will fail their EOG. You want to make sure your kids go to bed early so they can stay awake and think better they might pass the [S] EOG. [G]If they go to bed and go to school later they can do better in class and they could pass their grade.

Comments

The helping verb "didn't" requires a different form of the verb "went" from the one you've used. For example, instead of saying "I haven't finish the game," say "I haven't FINISHED the game." Instead of saying, "I didn't wrote the message," say "I didn't WRITE the message."

〈　〉　　　　　　　　　　　Close

Kids need r[...] They need to know about life things with math, reading, [S]S.S, ETC. They need learn mo[...] to be smarter say they have kids they can with their kids if they are having trouble[...] falut if they don't go to bed early and they are sleepy. They should learn to pay attenti[...] eir fault not getting enough sleep.

[G]In conclu[...] lementary go to school later why shouldn't the middle school and the high school. [S]Their[...] more and pay attention in class. [S]Their will be more kids passing the [S] EOG and the [S]EOC. It's not fair we have to start school earlier and get out later while [G][S]their are Elementary starts school later get out sooner. [G]You was a kid once you know how it feels being sleepy and you have no idea what they are talking [G]about or you haven't learn anything. [G]If kids got more sleep they will make a better grade because they are wide awake.

Submissions

→ Latest Draft
5th Draft
4th Draft
3rd Draft
2nd Draft
1st Draft

✎ Revise

24 submissions left

T Format

Peer Review

✉ Request

Figure 2. Student essay annotated with spelling and grammar feedback.

Style

★ ★ ★ ★ ☆
You scored 3.8 out of 5

Evaluation

- The reader gets a sense of who you are through your writing.
- Your personal view is present throughout the entire essay.
- Your use of language is natural and helps bring the topic to life.
- Your writing fits the task, purpose, and readers.
- Your writing seems to show what you really think.

Lessons

- Enhancing Style and Effect   • Understanding Audience

Word Choice

★ ★ ★ ⯪ ☆
You scored 3.6 out of 5

Evaluation

- Your word choice makes your writing interesting but easy to understand.
- Your words are natural and help to make the essay clear and believable.
- Your words are used in a generally interesting way.

Feedback

- Check your writing for places you can choose words that are more specific to make your writing more informative. Try adding strong verbs, specific nouns, adjectives, and adverbs.
- You have used some words incorrectly. Check for homophones, words that sound the same but are spelled differently and have different meanings, or for words that are often confused with other words.

Lessons

- Appositives

Figure 3. Writing analysis with evaluation and feedback for each trait.

Category

Development of Ideas

Organization

Style

Word Choice

Sentence Structure

Conventions

Grammar Exercises

Passage Exercises

Difficulty Level

All Levels

Beginner

Intermediate

Advanced

Using Self-Statements to Improve Writing

🕐 Time: 10 minutes   Level: Intermediate   🔊 Read-aloud available

Expert writers talk to themselves as they write. They use self-statements to help them get started, evaluate their progress, and encourage them to keep going. In this lesson, you will see examples of self-statements and how they can be used to improve your writing.

Strategies for Planning and Writing an Argumentative Essay

🕐 Time: 20 minutes   Level: Intermediate   🔊 Read-aloud available

This lesson introduces two powerful strategies to use when planning and writing an argumentative essay: STOP and DARE. Join Mrs. Norris and her students to learn about these strategies and how you can use them with your next argumentative essay.

Text Evidence in Argumentative Essays

🕐 Time: 14 minutes   Level: Intermediate   🔊 Read-aloud available

Adding text evidence in the form of direct quotations or paraphrasing can make your argumentative essay stronger and more credible. In this lesson, you will learn how to select evidence and how to use it correctly.

Figure 4. Interactive lessons.

argumentative writing performance. The study involved two treatment conditions in which students used NC Write to compose essays, receive automated scores and feedback, revise essays based on feedback, and complete interactive lessons during an eight-week intervention. In one treatment condition teachers integrated NC Write into their traditional process writing instruction (NC + TRAD); in a second treatment condition teachers integrated NC Write into strategy instruction, specifically, Self-Regulated Strategy Development (SRSD) instruction (Harris & Graham, 1988), which was adapted to a lower-intensity format to support teacher implementation (NC + SRSD).

Students in the NC + TRAD condition were exposed to instruction that incorporated authentic writing opportunities and cycles of planning, drafting, and revision. Students in the NC + SRSD condition were exposed to explicit instruction on cognitive and metacognitive strategies for planning and writing argumentative text. Students in a third, comparison condition received traditional process writing instruction and did not use NC Write. Multilevel model results showed that NC + SRSD students produced the highest-quality essays at posttest; these essays were also longer and included more basic elements of argumentative essays than those produced by students in the other two conditions. Results also showed NC + TRAD students produced higher-quality essays at posttest than comparison students. Thus, in addition to research showing that MI Write facilitates teachers' provision of a greater proportion of feedback on higher-level writing skills (Wilson & Czik, 2016), supports improvements in students' writing motivation and self-efficacy (Wilson & Czik, 2016; Wilson & Roscoe, 2020), and scaffolds improvements in students' writing quality across successive drafts to an essay (Wilson, 2017; Wilson & Czik, 2016; Wilson et al., 2014), Palermo and Thomson (2018) showed that AWE could be combined with different instructional approaches and thereby increase instructional efficacy.

2.1 The Present Study

The present study leverages data collected during the Palermo and Thomson (2018) study to provide a more nuanced examination of how AWE can be integrated in two different teacher-directed instructional contexts, the related effects on students' growth in first-draft writing quality across several essays over time, and teachers' and students' perceptions of AWE in those contexts. These features of the present study are unique and important. No prior study of AWE usage in ELA classrooms has examined two different models of integrating AWE with teacher-led instruction, yet such research is essential if educators are to understand how best to leverage the affordances of AWE for maximal instructional benefit. In addition, prior studies of AWE efficacy have rarely examined growth in first-draft writing quality across multiple essays, focusing instead on growth across revisions of a single essay or growth from pretest to posttest. Examining growth in

first-draft writing quality allows for testing whether prolonged exposure to instruction integrated with AWE feedback supports generalized improvements in independent writing performance. Finally, though teacher and student perceptions are important moderators of AWE efficacy (see Wilson, 2017), it is rare that their voices are included in research on AWE (c.f., Grimes & Warschauer, 2010; Warschauer & Grimes, 2008; Wilson & Roscoe, 2020).

Thus, the present study examined writing instruction with AWE with the goal of understanding how AWE might support the teaching and learning of writing. We employed an embedded quasi-experimental mixed methods design to investigate writing instruction supported by the AWE system NC Write. Two research questions were addressed: (1) What are students' first-draft writing performance (i.e., writing quality, essay length, and essay elements) growth trajectories when AWE is used within two instructional contexts: process writing instruction and strategy instruction? (2a) What are students' and teachers' experiences with and perceptions of the AWE system? (2b) What do these perceptions indicate regarding the affordances of AWE in the different instructional contexts?

3. Evidence of Effectiveness: A Mixed-Methods Examination of Writing Instruction with AWE

3.1 Methods

The present study followed an embedded quasi-experimental, after-intervention mixed methods design (Clark & Creswell, 2008). The intent of this design was to embed qualitative data collection within a quantitative experiment in order to supplement the quantitative results (Creswell, 2015).

Setting and participants

This study draws from the sample described in Palermo and Thomson (2018). The present study involved additional measures and different analyses, and addressed different research questions.

The teacher sample included 14 teachers in five districts who were assigned to integrate and implement NC Write within their traditional process writing instruction (NC + TRAD), or implement NC Write integrated with strategy instruction in the form of SRSD instruction (NC + SRSD). Students were recruited and assigned to conditions (within classes) by teacher. In this regard, the study employed a quasi-experimental design in which condition assignment was at the teacher level and outcomes were measured at the student level. All students in the classes of participating teachers completed study activities but only the data of those students who provided informed consent to participate were analyzed. At

the start of recruitment, total enrollment in participating teachers' classrooms was 1043 students. Of these students, 677 (65%) provided consent to participate. However, 118 consenting students were excluded from the sample due to incomplete pre-/posttests, insufficient composition during the intervention (<3 essays total), and/or changing class/teacher assignment over the course of treatment. The final sample comprised 559 students in grades 6 through 8. Table 1 presents participant demographics. As the two conditions were not equivalent in all respects, differing in the proportion of Black students and students with disabilities (SWDs), we examine the impact of these variables in the analyses.

Students who provided consent to be interviewed were further sampled to collect qualitative data. The interview sample comprised 30 students, 15 per condition. In addition, 12 teachers were interviewed, six per condition.

Time frame and conditions

The intervention was conducted in the Spring of 2016. None of the participating teachers or students had experience with NC Write prior to the Spring of 2016. In the first week, students composed pretest essays. Then, for a period of eight weeks, students received traditional writing instruction or SRSD writing instruction from their teachers and used NC Write.

Students in both conditions used NC Write to write and receive feedback for five unique essays (excluding pretest/posttest essays), revise each essay using feedback, and complete a total of six interactive lessons. Teachers in both conditions provided writing instruction during two 45-minute classes per week. In two districts, unanticipated events required an additional two weeks for teachers in both conditions to complete instruction. Total instructional time was 12 hours in both conditions. Students composed posttest essays one week after the intervention.

While there was some variability between and within districts in technological resources, on average all participating schools provided at least one digital learning device for each student (see Table 1). Most schools also had a bring-your-own-device policy. Thus, students had sufficient access to technology to use NC Write as part of writing instruction, either via a school device (e.g., Chromebook) or via a personal device.

3.2 Measures

Writing prompts

Argumentative writing prompts were used to assess students' writing performance. All prompts were reviewed by subject-matter experts. Prompts addressed a variety of topics relevant to middle school students. Some prompts

included sources, such as a brief article and/or video. At pre- and posttest, prompts were assigned to students using a counterbalanced design. During the intervention, students were provided a choice of two prompts for each essay. As our primary interest was to examine students' growth in independent, overall writing performance, an outcome rarely examined in prior research, we analyzed growth in the quality of students' first drafts across essays. One essay was written collaboratively and excluded from analysis.

Table 1. Participant demographics

Variable	NC + TRAD	NC + SRSD	Difference test
Students (n)	272	287	
Districts (n)	3	4	
Schools (n)	3	5	
Teachers (n)	6	8	
Grade (n)			
6	41	113	
7	160	28	
8	71	146	
Male (%)	48.71	45.10	$\chi^2_{(1)} = 0.726, p = .394$
Race (%)			
White	78.71	50.74	$\chi^2_{(1)} = 54.530, p < .001$
Hispanic or Latino	12.93	12.59	$\chi^2_{(1)} = 0.013, p = .908$
Black or African American	4.18	32.96	$\chi^2_{(1)} = 72.403, p < .001$
Asian	1.14	0.37	$\chi^2_{(1)} = 1.061, p = .303$
American Indian or Alaska Native	0.38	0.37	$\chi^2_{(1)} = 0.000, p = .985$
Native Hawaiian or Pacific Islander	0.38	-	$\chi^2_{(1)} = 1.029, p = .311$
Two or More Races	2.28	2.96	$\chi^2_{(1)} = 0.242, p = .623$
Free or Reduced Price Lunch[a] (%)	60.33	59.96	
Limited English Proficiency (%)	3.06	5.71	$\chi^2_{(1)} = 2.062, p = .151$
Students with Disabilities (%)	5.68	14.12	$\chi^2_{(1)} = 9.529, p = .002$
Age[b] (M, SD)	155.51	156.42	$F(1,557) = 0.981, p = .322$
	(9.11)	(12.31)	
Previous ELA achievement[c] (M, SD)	455.67	454.53	$F(1,557) = 1.311, p = .253$
	(10.53)	(11.61)	
Students per digital learning device[a] (M)	0.73	1.01	

[a] Estimates based on school-level data.

[b] Age in months at time of pretest.

[c] Based on scale score (range: 423–484) from the previous year's end-of-grade assessment.

Writing quality

The writing quality of each essay was measured by the PEG total essay score (range = 6–30). We opted to analyze the total essay score rather than trait scores for two reasons. One, our quantitative research question addressed students' growth trajectories in overall writing quality, which is the construct that the PEG total essay score intends to measure. Two, individual trait scores have been found to be highly correlated (Wilson et al., 2014).

Essay length

The length of each essay was measured in number of words. All written words, regardless of spelling, were included in this calculation. Microsoft Excel was used to calculate essay length.

Essay elements

Each essay was evaluated for the basic elements of argumentative writing, using procedures described by Scardamalia, Bereiter, and Goelman (1982). Elements evaluated included claim, supporting reasons, elaborations, counterclaims, and conclusion. Essays were assigned one point for each element present, or, in the case of supporting reasons, elaborations, and counterclaims, one point for each separate and unique example included.

Professional raters were enlisted to score essays for the basic elements of argumentative writing. Raters were trained by a scoring director and team leader, using a scoring guide and set of anchor essays. Following, raters evaluated and discussed two, 10-essay practice sets. Lastly, raters were exposed to challenging-to-score essays in the form of an 8-essay supplemental training set. Following one and a half days of training, raters scored essays, using training materials, particularly anchor essays, to ground their scoring decisions in the scoring criteria. The scoring director and team leader spot-checked raters' scores during scoring in order to evaluate accuracy and drift and provide feedback to raters. A random 15% of essays were scored a second time by the scoring director or team leader to assess inter-rater reliability. Reliability was 99% exact agreement for claims ($r = .924$), 78% for supporting reasons ($r = .923$), 64% for elaborations ($r = .917$), 96% for counter-claims ($r = .874$), and 96% for conclusions ($r = .928$).

Students' and teachers' experiences with and perceptions of the AWE system

Students and teachers participated in one-on-one, semi-structured interviews based on open-ended questions to investigate their experiences with and perceptions of the AWE system and what these perceptions indicated regarding the affordances of AWE in the different instructional contexts. We had no a priori assumptions about user experiences and perceptions (due in part to limited and

dated prior research related to user experiences with and perceptions of AWE in implementation contexts), thus a broad and common set of questions related to NC Write were posed to students and teachers in both conditions. Students were asked questions including whether and in which ways NC Write helped them improve their writing, whether the automated feedback helped them improve their writing, what they learned from using the program, and what they would change about NC Write. Teachers were asked similar questions including whether and in which ways NC Write helped their students improve their writing, their opinions regarding the automated feedback, and what they would change about NC Write. Interviews were recorded digitally and transcribed verbatim.

We employed qualitative data analysis strategies borrowed from grounded theory (Strauss & Corbin, 1998) to analyze interview data. Analysis was conducted using QSR International's NVivo 10 software. Two individuals experienced in qualitative data analysis coded data in an iterative and recursive process. Stages of coding followed the progression recommended by Corbin and Strauss (2015). First, interview responses were grouped by condition and open coded to determine initial concepts. Next, we used comparative analysis to establish the associated dimensions and properties of each concept. Following, we coded responses for context, process, and integration of categories. We validated the final coding scheme against negative cases identified during coding. The final coding scheme appears in Table 2.

3.3 Procedures

General instructional procedures

Before the study began, all teachers participated in a walk-through of NC Write to develop familiarity with the basic features of the program. Teachers also received their account information and had the opportunity to explore the software prior to instruction.

Teachers were provided with all study materials needed to conduct the intervention. Each teacher received detailed directions outlining the study expectations and treatment procedures, pretest and posttest instructions, and all student materials needed during the intervention. Teachers also received a schedule overview illustrating NC Write activities (i.e., interactive lessons, essay planning and writing, essay revising) by class, as well as lesson plans for each class.

Due to previous instances of students and teachers placing undue faith in AWE (Grimes & Warschauer, 2010), teachers were given a script to follow in order to provide students with consistent background information about PEG and automated essay scoring. Students were informed that PEG does not read and understand essays in the way that humans do, but knows how to find

characteristics of good writing from training that involved lots of essays that were read and scored by humans. Students were reminded that PEG could only accurately score "good faith" essays and that if they submitted plagiarized text or otherwise tried to deceive PEG their scores may not accurately reflect the quality of their writing.

Table 2. Coding Scheme

Concept	Dimensions/Properties	Description
1. Benefits of NC Write	1.1 Efficiency 1.2 Evidence of growth 1.3 Provides a structure for writing 1.4 Relevance 1.5 Supports differentiation 1.6 Supports intrinsic motivation 1.7 Supports writing instruction	Perceptions of benefits of NC Write for learning and teaching
2. Change about NC Write	2.1 Opinions	Aspects of NC Write users felt should be improved, added, or removed
3. Writing quality feedback	3.1 How feedback supported improvement 3.2 How students used feedback 3.3 Teacher assistance with interpretation 3.4 Criticisms and other limitations of feedback	Perceptions, implementation, and use of automated feedback
4. NC Write Lessons	4.1 How supported improvement 4.2 Criticisms and limitations	Perceptions of interactive lessons
5. Prompts	5.1 Perceptions	Perceptions of intervention prompts
6. Graphic organizers	6.1 How supported improvement 6.2 Criticisms and limitations	Perceptions of graphic organizers in NC Write
7. Explanation for growth trajectories	7.1 Ceiling on improvement	Factors explaining growth trajectories
8. Implementation challenges	8.1 Practice 8.2 Intervention timing	Challenges associated with implementing NC Write and/or intervention

Teachers in the SRSD condition were further trained to implement strategy instruction in the form of SRSD for writing. Training was conducted on-site at teachers' respective schools and lasted for one hour. First, teachers were provided

with background information about the SRSD model. Training comprised background on the theoretical base of SRSD; the rationale for SRSD, and the stages of SRSD writing instruction. During training teachers reviewed the provided SRSD lesson plans and associated student materials. It was emphasized that lesson plans were intended to serve as a framework for instruction and should be adapted as needed in order to differentiate instruction based on learners' needs.

NC + TRAD instructional procedures

Teachers in the NC + TRAD condition use a traditional process approach to writing instruction. Process writing instruction is defined by extended, authentic writing opportunities; cycles of planning, writing, and revising; a supportive, interactive, writing environment; individualized instruction; student ownership of writing; and self-reflection and evaluation (Graham, McKeown, Kiuhara, & Harris, 2012; Graham & Perin, 2007). This instruction addressed the basic structure of argumentative essays, planning and organizing strategies, and a variety of relevant writing skills. Students in the NC + TRAD condition used NC Write to access interactive lessons that focused on the development of ideas, organization, word choice, sentence structure, and conventions; to write argumentative essays; and to revise essays using feedback provided by the program.

NC + TRAD teachers' process writing procedures were described in detail in Palermo and Thomson (2018). In brief, results of a survey of writing instructional practices (Gilbert & Graham, 2010) confirmed that teachers used a process approach to writing instruction that included writing skills instruction. The evidence-based practices that NC + TRAD teachers reported using most frequently included direct instruction of skills, summarization instruction, writing as a learning tool, and paragraph writing. Commonly reported, but less frequently used practices included spelling, sentence combining, and inquiry/research.

NC + TRAD teachers' interview results provided additional insights into their practices. Five of the six teachers shared that writing instruction had become less of a priority since the elimination of the North Carolina General Writing Assessments. Teachers described using a variety of tactics such as increasing the integration of reading and writing, incorporating elements of argumentative writing into research papers, and utilizing abbreviated writing sessions as a result. A number of the teachers espoused a five-paragraph essay structure, acknowledging this approach as formulaic. NC + TRAD teachers primarily incorporated writing skills into their process writing instruction through mini-lessons several times a week.

NC + TRAD lessons provided teachers with a general framework to support teachers' integration of NC Write within their traditional process writing instruction, and included the following guidance. One, provide students with

background information about PEG and automated essay scoring using the supplied script. Two, give students the opportunity to compose a unique essay in NC Write approximately every other week (i.e., five total, not including pretest and posttest), providing a choice of prompts. Three, support students in selecting appropriate graphic organizers in NC Write for planning and pre-writing. Four, provide each student with the opportunity to revise each essay using automated feedback. Emphasize the importance of using the feedback to make substantive revisions that improve the structure of the essay. Five, assign each student an NC Write lesson approximately every other week (i.e., six total). Recommended lessons for NC + TRAD students featured topics such as understanding audience, elaboration in essays, and sentence variety.

In sum, NC + TRAD teachers used a process writing approach that included AWE to teach students the basic structure of argumentative essays and a variety of writing skills. While process writing instruction shared some similarities to SRSD writing instruction, it differed in three important ways: (1) NC + TRAD teachers did not teach strategies for planning, writing, and revising argumentative essays to the extent these strategies were taught by NC + SRSD teachers; (2) NC + TRAD teachers did not explicitly teach self-regulation procedures; and (3) NC + TRAD instruction was not criterion based and/or instructionally responsive to the same extent as NC + SRSD instruction.

NC + SRSD instructional procedures

Teachers in the NC + SRSD condition implemented SRSD writing instruction. SRSD writing instruction addressed strategies for planning and writing argumentative essays; the knowledge, skills, and self-regulatory procedures to apply strategies; and positive beliefs about writing. Students in the NC + SRSD condition used NC Write to access interactive lessons that emphasized self-regulation procedures, to write argumentative essays, and to revise essays using feedback provided by the program.

NC + SRSD teachers implemented six recursive stages of SRSD writing instruction. This instruction aims to provide students with general and genre-specific writing strategies, the knowledge to apply strategies, and self-regulatory knowledge and abilities to effectively apply strategies and manage the writing task (Graham, Harris, & McKeown, 2013; Harris & Graham, 2016). Students are also taught a short mnemonic for each strategy to help them recall the cognitive processes associated with strategy application (Graham & Harris, 2018). The stages of instruction support a gradual release of responsibility as students develop proficiency with strategy use, as well as encourage maintenance and generalization of strategies.

The overall effectiveness of SRSD on students' writing performance has been verified in several meta-analyses (i.e., Graham, 2006, Graham & Harris, 2003;

Graham et al., 2012; Graham & Perin, 2007). However, most SRSD interventions described in the literature have been facilitated by researchers or tutors, conducted in small group or one-on-one settings, and involved high-intensity treatment (i.e., 20–45 minute lessons administered three to five days per week). More recently, a limited number of studies have examined classwide teacher implementation of SRSD following substantial practice-based professional development (e.g., Festas et al., 2015; McKeown et al., 2016, 2018). For the purpose of the present study, we adapted the SRSD model to a lower-intensity format with the goal of supporting teacher implementation and reducing professional development.

As the present study involved argumentative writing, students were taught the **STOP** planning strategy (**S** = Suspend judgment, **T** = Take a side, **O** = Organize ideas, **P** = Plan more as you write). One intent of this strategy is to encourage students to choose the side of an argument they could best defend considering claims and potential counterclaims. This strategy also encourages students to consider how to organize this information to make a compelling argument, and to continue to plan, adjust, and revise during the writing process. Students were further taught the genre-specific **DARE** strategy (**D** = Develop your claim, **A** = Add supporting reasons and evidence, **R** = Reject counterclaims, **E** = End with a conclusion). The purpose of this strategy is to help students include all of the basic elements of argumentative essays in their compositions.

The specific steps teachers used to implement the six stages of SRSD writing instruction and teach students the planning and writing strategies were described in detail in Palermo and Thomson (2018). In brief, in the first stage, *Develop background knowledge,* students were provided with the necessary knowledge and vocabulary to begin applying the strategies. Teachers introduced the planning and writing strategies, which students practiced recalling and applying. This scaffolded practice continued throughout the remaining stages of instruction. In the second stage, *Discuss it,* teachers and students continued to discuss the strategies, in particular how the basic elements of argumentative essays make an essay effective and convincing.

In the third stage, *Model it,* teachers demonstrated application of **STOP** and **DARE** by modeling composition of an argumentative essay. While composing, teachers also modeled self-regulation procedures including goal setting, self-instruction, and self-evaluation. The fourth stage, *Memorize it*, comprised various activities to support students' recall and application of the two strategies. Teachers assessed students' knowledge of the strategies regularly, differentiating support and practice activities until each student had internalized the strategies.

In the fifth stage, *Support it*, students' application of the planning and writing strategies and self-regulation procedures were supported by both teachers (via collaborative essay writing activities) and peers (via small-group revision

activities). The bulk of students' essay composition and revision occurred during this stage; teachers faded support as students developed proficiency using the **STOP** and **DARE** strategies and the self-regulation procedures. The final stage, *Independent performance,* was reached when students demonstrated the ability to plan and write an effective argumentative essay independently, i.e., in the absence of teacher support and/or PEG feedback.

NC + SRSD lessons included guidance for teachers to integrate NC Write with the strategy instruction. This guidance included the same recommendations made to NC + TRAD teachers, namely: provide background information about PEG and automated scoring; have students compose a total of five unique essays in NC Write; ensure students select appropriate graphic organizers for prewriting; provide time for students to revise each essay; and assign six NC Write lessons. What distinguished the integration of NC Write with strategy instruction was the alignment of NC Write lessons, composing, and revising activities with the stages and content of teacher-led SRSD writing instruction. For example, NC + SRSD teachers assigned specific NC Write lessons at particular stages of instruction to reinforce the strategies for planning and writing and self-regulation procedures introduced to students during teacher-led instruction. NC Write lessons assigned to NC + SRSD students featured topics such as strategies for planning and writing an argumentative essay, using self-statements to improve writing, and improving essays with transitions. Students' composition and revision activities in NC Write peaked during the fifth stage of instruction (*Support it*), which required the greatest amount of writing practice as students developed writing proficiency and teachers faded their support.

AWE exposure

In the absence of the ability to conduct direct observations of classroom instruction in each condition, we elected to analyze NC Write log-files to evaluate each student's exposure to AWE during writing instruction. Using those log-files, we produced a measure that represented each student's AWE exposure based on the primary treatment activities. This measure was a count of the total treatment activities each student completed during the intervention, including NC Write lessons, first drafts, and revisions, and had a maximum of 17. This count excluded any essays submitted with a high proportion of plagiarized text and lessons completed quickly (i.e., <2 minutes). NC + SRSD students (M = 13.73, SD = 2.73) completed more total treatment activities than NC + TRAD students (M = 12.86, SD = 2.48); t(557) = 3.93, p < .01). The AWE exposure measure was included in subsequent analyses using grand mean centering.

3.4 Analyses

Multilevel modeling was used to conduct quantitative analyses. This framework was capable of accommodating the complete dataset that included unbalanced data and non-independent observations. All analyses were conducted using MLwiN v3.02 (Charlton, Rasbash, Browne, Healy, & Cameron, 2017).

To examine students' writing performance growth trajectories, a series of three-level models were specified to incorporate all first-draft essays written by students in the NC + TRAD and NC + SRSD conditions. During the course of the intervention students completed approximately six first-draft essays ($M = 6.31$, $SD = .92$). Preliminary review of individual students' writing performance outcomes (i.e., writing quality, essay length, and essay elements) showed variability in outcomes at pretest (i.e., the first essay) and variability in change over time. Specifically, many students demonstrated curvilinear growth in outcomes over time in the form of fairly rapid improvement in performance following the start of treatment but slower growth as the intervention progressed. Based on previous research examining writing performance growth trajectories associated with AWE (Wilson, 2017; Wilson et al., 2014), several growth models were tested: a linear model, a logarithmic model using the natural log of the time variable, and a polynomial, quadratic growth model using the variables *Time* (a count of each unique first-draft essay completed used to describe instantaneous rate of change) and *Time²* (a square of the *Time* variable used to describe the deceleration in students' growth curves). Both variables were centered such that 0 represented students' pretest writing performance.

For all analyses, an unconditional model (Model 1) was first specified to calculate intra-class correlations and ensure there was sufficient variance at all levels to justify a three-level model. Results of all unconditional models indicated there was sufficient variability at all levels to warrant a three-level model. The following quadratic growth model (Model 2) was then used to examine students' writing performance growth trajectories:

$$y_{ijk} = \beta_{0ijk} + \beta_1 \text{Time}_{ijk} + \beta_2 \text{Time}^2_{ijk} + v_{0k} + v_{1k}\text{Time}_{ijk} + v_{2k}\text{Time}^2_{ijk} + u_{0jk} + u_{1k}\text{Time}_{ijk} + u_{2k}\text{Time}^2_{ijk} + e_{ijk}$$

In this model, y_{ijk} represents the predicted writing performance outcome (i.e., writing quality, essay length, essay elements) for student i with teacher j at time k. β_{0ijk} is the initial (pretest) status of student i with teacher j for the respective predicted outcome, $\beta_1\text{Time}_{ijk}$ represents the instantaneous growth rate for student ij at time k, $\beta_2\text{Time}^2_{ijk}$ represents the curvature (i.e., deceleration) in the growth rate for student ij, v_{0k} represents between-teacher variance, u_{0jk} represents between-student variance, and e_{ijk} represents within-student variance. Random

effects are assumed normally distributed with means of zero and constant variances σ^2.

Finally, a conditional quadratic growth model (Model 3) was specified in the form:

$$y_{ijk} = \beta_{0ijk} + \beta_1 \text{Time}_{ijk} + \beta_2 \text{Time}^2_{ijk} + \beta_3 \text{Exposure}_{jk} + \beta_4 \text{SRSD}_k + \beta_5 \text{SRSD*Time}_{ijk} + \beta_6 \text{SRSD*Time}^2_{ijk} + v_{0k} + v_{1k} \text{Time}_{ijk} + v_{2k} \text{Time}^2_{ijk} + u_{0jk} + u_{1k} \text{Time}_{ijk} + u_{2k} \text{Time}^2_{ijk} + e_{ijk}$$

This model expands on the previous quadratic growth model (i.e., Model 2), adding a control for AWE exposure ($\beta_3 \text{Exposure}_{jk}$) and predictors to test for between-condition differences in initial status ($\beta_4 \text{SRSD}_k$), growth rate ($\beta_5 \text{SRSD*Time}_{ijk}$), and deceleration ($\beta_6 \text{SRSD*Time}^2_{ijk}$) for each writing performance outcome.

Fit of all models was examined using negative log likelihood (–2LL). Smaller values indicate better fit provided a significant chi-square difference test based on the number of estimated parameters.

3.5 Results

What are students' first-draft writing performance (i.e., writing quality, essay length, and essay elements) growth trajectories when AWE is used within two instructional contexts: process writing instruction and strategy instruction?

Writing quality

Table 3 presents results of writing quality models. Unconditional model (Model 1) random effects showed that 49% of the variability in writing quality fell within students, while 27% fell between students within teachers and 24% fell between teachers. The quadratic growth model (Model 2) provided significantly better fit to the data than the unconditional model [$\chi^2_{(12)}$=1095.52, p <.001]. Results of the quadratic model showed, accounting for the clustered data structure, that students improved the quality of their first drafts by 1.6 points each subsequent essay, with a deceleration rate of –0.2 points. Tests of demographic × condition interactions produced non-significant results for initial status and rate of change for Black students (p = .47 and .09, respectively), though NC + SRSD Black students exhibited slightly slower deceleration rates than NC + SRSD non-Black students (p = .01).

No significant differences were observed for initial status, rate of change, or deceleration of SWDs (p = .38, .56, and .90, respectively); for parsimony these variables were omitted from the final model. The final conditional model (Model 3), which controlled for AWE exposure, showed that initially, NC + SRSD students produced first drafts that were of a lower quality (by 2.4 points) than NC + TRAD students. Over time, NC + SRSD students demonstrated nonsignificant differences

in writing quality growth rate (by 0.68 points) and deceleration (by −0.07 points) compared to NC + TRAD students. Figure 5 presents the predicted growth trajectories for both groups of students, separately for Black and non-Black students.

Essay length

Table 4 presents results of essay length models. Unconditional model (Model 1) random effects showed, of the total variability in essay length, 47% fell within students, while 27% fell between students within teachers and 26% fell between teachers. The quadratic growth model (Model 2) better fit the data [$\chi^2_{(12)}$=1061.69, p <.001]. Model results showed, after accounting for the clustered data structure, students increased the length of their first-drafts approximately 45 words each subsequent essay, with a deceleration rate of −6 words. Tests of demographic × condition interactions for initial status, rate of change, and deceleration produced non-significant results for Black students (p = .15, .16, and .22, respectively) and SWDs (p = .89, .97, and .60, respectively) and these variables were omitted from the final model. The final conditional model (Model 3), controlling for AWE exposure, showed that NC + SRSD students initially produced shorter first drafts than NC + TRAD students (by 102 words). Over subsequent first-draft essays, NC + SRSD students demonstrated nonsignificant differences in essay length growth rate (by 19 words) and deceleration (by 2 words) compared to NC + TRAD students. Figure 6 depicts these predicted growth trajectories.

Essay elements

Table 5 presents results of essay element models. Unconditional model (Model 1) random effects indicated that of the total variability in the number of basic elements of argumentative essays, 49% fell within students, while 27% fell between students within teachers and 24% fell between teachers. The quadratic growth model (Model 2) provided significantly better fit to the data than the unconditional model [$\chi^2_{(12)}$=649.59, p <.001]. Model results showed, after accounting for the clustered data structure, students increased the number of elements in their first drafts over time by approximately 1.6 elements each subsequent essay, with a deceleration rate of −0.2 elements. Tests of demographic × condition interactions for initial status, rate of change, and deceleration produced non-significant results for Black students (p = .49, .96, and .64, respectively) and SWDs (p = .35, .32, and .24, respectively) and these variables were omitted from the final model.

Table 3. Unstandardized Coefficients (and Standard Errors) of Multilevel Growth Models of Essay Quality

Parameter	Model 1: Unconditional model			Model 2: Quadratic growth model			Model 3: Conditional model		
	Est.	SE	p	Est.	SE	p	Est.	SE	p
Fixed Effects									
Writing quality									
Initial status, β_0	14.11	0.57	<.001	12.05	0.68	<.001	13.54	0.84	<.001
AWE exposure, β_3							0.37	0.06	<.001
NC + SRSD, β_4							−2.43	1.12	.03
Rate of change									
Intercept, β_1				1.55	0.33	<.001	1.18	0.47	.01
NC + SRSD, β_5							0.68	0.63	.28
Deceleration									
Intercept, β_2				−0.21	0.05	<.001	−0.17	0.07	.01
NC + SRSD, β_6							−0.07	0.09	.43
Random Effects									
Level 1 (time, σ^2_{e0})	8.68	0.25		5.12	0.17		5.12	0.17	
Level 2 (students)									
Initial status (σ^2_{u0})	4.86	0.40		4.33	0.56		4.04	0.54	
Rate of change (σ^2_{u1})				0.22	0.14		0.21	0.14	
Deceleration (σ^2_{u2})				0.00	0.00		0.00	0.00	
Level 3 (teachers)									
Initial status (σ^2_{v0})	4.26	1.70		6.12	2.44		3.99	1.63	
Rate of change (σ^2_{v1})				1.40	0.57		1.28	0.52	
Deceleration (σ^2_{v2})				0.03	0.01		0.03	0.01	
Goodness of Fit									
Deviance (-2LL)	15877.16			14781.64			14738.73		
Difference Test				$\chi^2_{12}=1095.52$		<.001	$\chi^2_{4}=42.91$		<.001

Note. Est. = Unstandardized parameter estimate, SE = standard error, p = p-value. *P*-values are not provided for random effects as the Wald test for these parameters is only approximate.

Table 4. Unstandardized Coefficients (and Standard Errors) of Multilevel Growth Models of Essay Length

Parameter	Model 1: Unconditional model			Model 2: Quadratic growth model			Model 3: Conditional model		
	Est.	SE	p	Est.	SE	p	Est.	SE	p
Fixed Effects									
Essay length									
Initial status, β_0	267.09	22.30	<.001	204.42	26.46	<.001	265.35	32.70	<.001
AWE exposure, β_3							8.84	2.05	<.001
NC + SRSD, β_4							−102.48	43.63	.03
Rate of change									
Intercept, β_1				45.46	10.99	<.001	35.44	15.99	.03
NC + SRSD, β_5							18.73	21.44	.38
Deceleration									
Intercept, β_2				−5.96	1.63	<.001	−4.99	2.40	.04
NC + SRSD, β_6							−1.95	3.22	.54
Random Effects									
Level 1 (time, σ_{e0}^2)	11998.11	342.60		7037.84	245.10		7036.77	245.32	
Level 2 (students)									
Initial status (σ_{u0}^2)	6749.61	549.91		3646.38	649.10		3498.87	640.83	
Rate of change (σ_{u1}^2)				786.87	254.44		788.31	255.42	
Deceleration (σ_{u2}^2)				19.01	6.31		19.04	6.35	
Level 3 (teachers)									
Initial status (σ_{v0}^2)	6609.15	2627.67		9397.93	3700.49		6175.28	2476.43	
Rate of change (σ_{v1}^2)				1522.54	636.77		1420.26	597.81	
Deceleration (σ_{v2}^2)				32.07	13.90		30.84	13.43	
Goodness of Fit									
Deviance (-2LL)	37646.10			36584.41			36562.13		
Difference Test				$\chi_{12}^2 = 1061.69$		<.001	$\chi_4^2 = 22.28$		<.001

Note. Est. = Unstandardized parameter estimate, SE = standard error, p = p-value. P-values are not provided for random effects as the Wald test for these parameters is only approximate.

Table 5. Unstandardized Coefficients (and Standard Errors) of Multilevel Growth Models of Essay Elements

Parameter	Model 1: Unconditional model			Model 2: Quadratic growth model			Model 3: Conditional model		
	Est.	SE	p	Est.	SE	p	Est.	SE	p
Fixed Effects									
Essay elements									
Initial status, β_0	9.28	0.58	<.001	7.28	0.72	<.001	8.90	0.89	<.001
AWE exposure, β_3							0.25	0.07	<.001
NC + SRSD, β_4							−2.71	1.20	.02
Rate of change									
Intercept, β_1				1.63	0.34	<.001	1.33	0.50	<.01
NC + SRSD, β_5							0.55	0.67	.41
Deceleration									
Intercept, β_2				−0.24	0.05	<.001	−0.20	0.08	.01
NC + SRSD, β_6							−0.06	0.11	.59
Random Effects									
Level 1 (time, σ_{e0}^2)	11.83	0.34		8.81	0.25		8.81	0.25	
Level 2 (students)									
Initial status (σ_{u0}^2)	6.29	0.52		6.85	0.52		6.73	0.51	
Rate of change (σ_{u1}^2)				0.00	0.00		0.00	0.00	
Deceleration (σ_{u2}^2)				0.00	0.00		0.00	0.00	
Level 3 (teachers)									
Initial status (σ_{v0}^2)	4.36	1.76		6.67	2.73		4.44	1.88	
Rate of change (σ_{v1}^2)				1.46	0.61		1.40	0.59	
Deceleration (σ_{v2}^2)				0.03	0.01		0.03	0.01	
Goodness of Fit									
Deviance (-2LL)	16775.71			16126.12			16109.57		
Difference Test				$\chi_{12}^2=649.59$		<.001	$\chi_{4}^2=16.55$		<.01

Note. Est. = Unstandardized parameter estimate, SE = standard error, p = p-value. P-values are not provided for random effects as the Wald test for these parameters is only approximate.
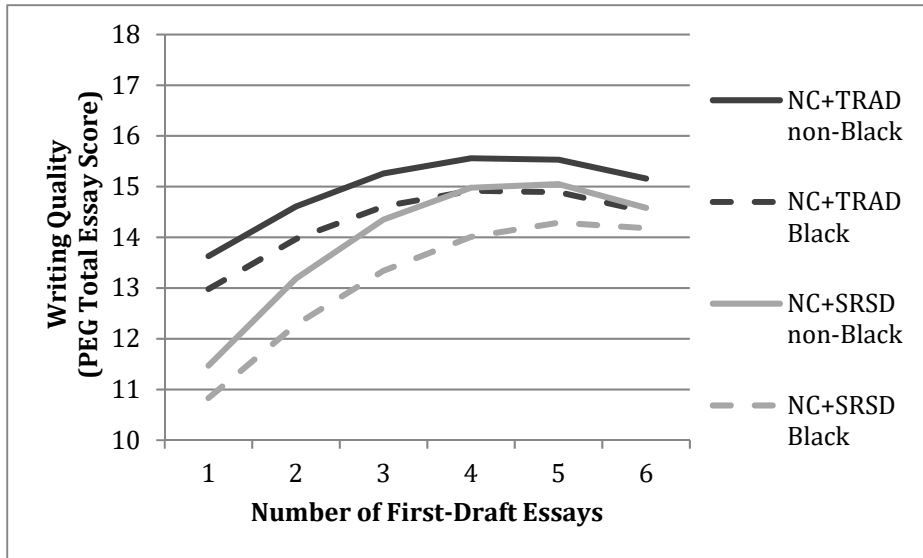
Figure 5. Growth in writing quality by treatment condition using restricted y-axis for illustration purposes (PEG total essay score range 6–30).
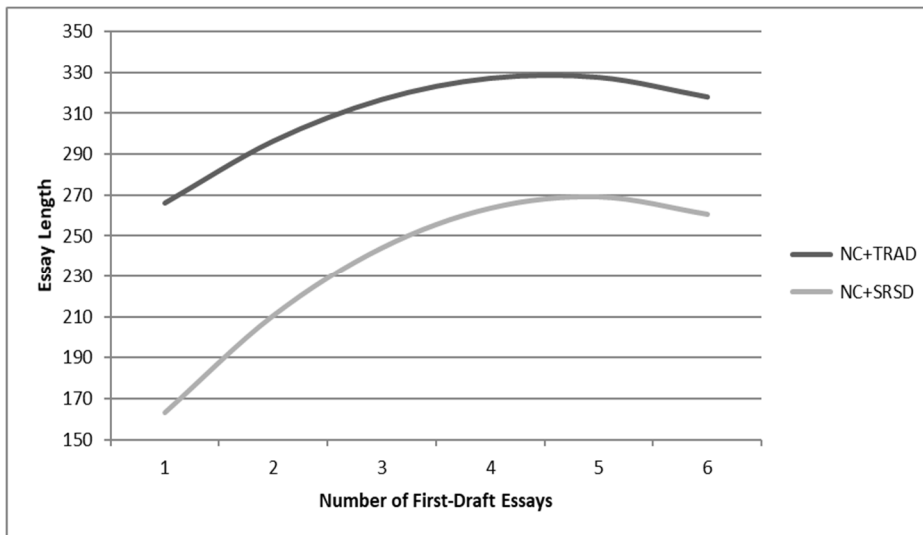

Figure 6. Growth in essay length by treatment condition using restricted y-axis for illustration purposes. Essay length.

The final conditional model (Model 3), which controlled for AWE exposure, showed that initially, NC + SRSD included 2.7 fewer basic elements in their first drafts than NC + TRAD students. Over time, NC + SRSD students demonstrated nonsignificant differences in essay element growth rate (by 0.6 elements) and deceleration (by −0.06 elements) compared to NC + TRAD students. Figure 7 presents the predicted growth trajectories for both groups of students.
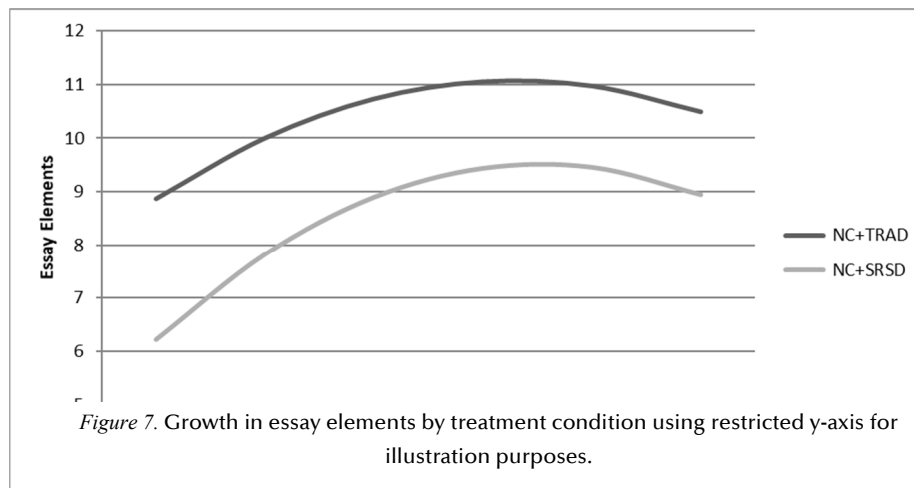
Figure 7. Growth in essay elements by treatment condition using restricted y-axis for illustration purposes.

What are students' and teachers' experiences with and perceptions of the AWE system and what do these perceptions indicate regarding the affordances of AWE in the different instructional contexts?

Students in both treatment conditions demonstrated writing performance growth across first-draft essays. Students' writing performance growth trajectories reflected initial, rapid growth that slowed over subsequent first-draft essays and eventually plateaued. Students' growth in writing performance followed this trajectory for all outcomes (i.e., writing quality, essay length, essay elements). Students in both treatment conditions exhibited comparable growth across first-draft essays, despite NC + SRSD students demonstrating lower writing performance initially. Findings suggest that prolonged exposure to teacher-led instruction integrated with AWE feedback was associated with generalized improvements in students' independent writing performance.

Qualitative data, in the form of students' and teachers' experiences with and perceptions of the AWE system, were collected to supplement the quantitative results and to probe aspects of the instruction and AWE feedback that may have contributed to the growth observed in the quantitative results. We borrowed qualitative data analysis strategies from grounded theory to determine concepts prevalent in interview data and establish the dimensions and properties of these concepts.

A major finding of the qualitative results was that teachers and students who used NC Write in process writing and strategy instruction contexts had similar experiences with and perceptions of the AWE system. This general finding, which

aligns with the quantitative results, suggested that NC Write provided homogeneous affordances—and limitations—across both instructional contexts. Therefore, in presenting qualitative results, we group results not by condition but instead organize findings by the primary concepts that emerged through analysis.

A framework for deliberate writing practice

In sum, qualitative analyses revealed that NC Write provided a framework for deliberate writing practice. Concepts that emerged from interview data helped to explicate how NC Write, when used to support writing instruction in two different instructional contexts, equally fostered students' growth in writing performance. The framework for deliberate writing practice is presented in Figure 8. This framework depicts writing as a cycle of learning, practice, and feedback. The arrows between each component are bidirectional, emphasizing the cyclical and non-sequential relations among components. Interview results indicated that NC Write provided a structure for writing in both instructional contexts that supported many of the key requirements of deliberate practice (Ericsson, 2006).
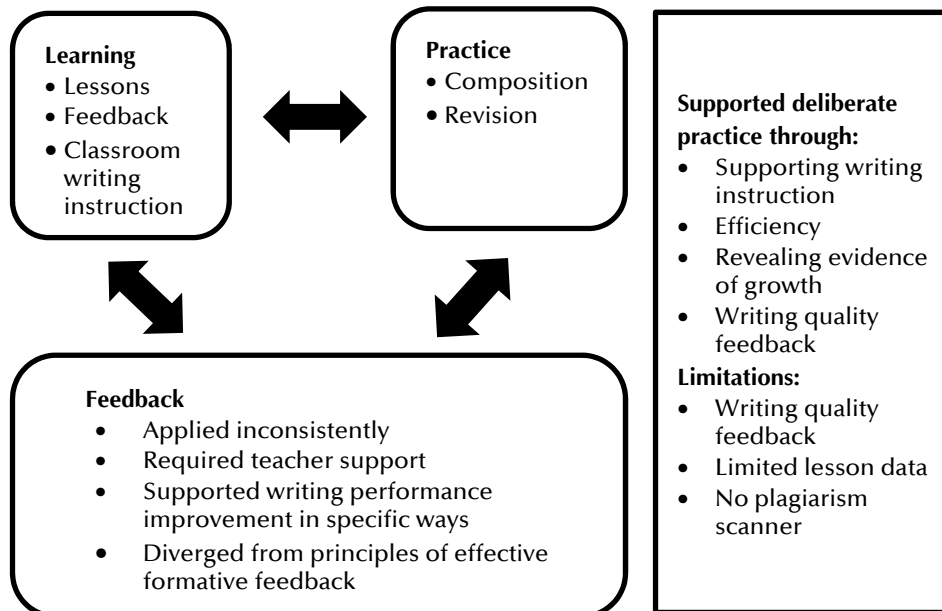

Figure 8. Framework for deliberate writing practice.

This core qualitative finding provides an explanation as to why the quantitative results did not show differential effects of instructional condition on students' growth trajectories: NC Write facilitated deliberate practice in both process writing and strategy instruction environments, making these instructional contexts more similar than they would be otherwise. In particular, the learning, practice, and feedback components of NC Write allowed students frequent and suitable practice opportunities and task-level performance feedback regarding the quality of their

writing. The interconnected nature of the components worked in concert to support students' growth in writing quality. Next, we discuss concepts that emerged from interview data and help to explain students' writing performance growth trajectories.

Supporting writing instruction

One concept that emerged from interview results was that NC Write supported teachers' process writing instruction and strategy instruction. NC Write was a tool for teachers to reinforce, and for students to apply, instructional strategies introduced in the classroom. For example, NC + SRSD students were introduced to the **STOP** and **DARE** strategies in a teacher-led lesson, and in a subsequent class students practiced the strategies in an interactive NC Write lesson and applied them by completing a graphic organizer and composing an essay in NC Write. One teacher explained how NC Write supported her SRSD instruction, sharing "the program provided us with examples that they were able to . . . really see how that structure worked together; that was very helpful and it was something that I didn't have to, myself, go and write."

A second property of NC Write supporting writing instruction was the alignment of the program with teachers' writing instruction. Teachers described how the program fit with their process approaches to writing and strategy instruction and was consistent with the writing standards. One NC + TRAD teacher explained, "I just feel like the NC Write program backs up my Common Core and what I'm already teaching. It's an additional tool for me that I think helps make my kids stronger writers." Teachers described making some adjustments to their instruction to increase this alignment. One NC + TRAD teacher shared:

> It was just taking the time to explain to them, and when I'm teaching, making sure that I'm using that same vocabulary so that they get used to it because I would not want to change the vocabulary that they used in the feedback because I liked it. I thought it was strong.

A third property of NC Write supporting writing instruction was that NC Write provided a structure for process writing instruction and practice. For some NC + TRAD teachers who did not feel well-prepared to teach writing, the program structure of lessons, composition and revision opportunities, and feedback was valuable. One teacher recounted:

> I'm not the best writing teacher and that helped me just have a basis of what to talk about and what to teach them. "OK guys, here's this prompt that's been given to us," and there would be lessons in here, "Let's talk about these things." I felt that NC Write helped me a lot with my instruction.

A final property of NC Write supporting writing instruction was that the program allowed renewed focus on writing in the classroom. In their interviews, the majority of teachers indicated that attention to writing instruction had atrophied in North Carolina in recent years, in part due to the elimination of the North Carolina General Writing Assessments. One NC + TRAD teacher explained:

> With technology and texting and all the ways kids communicate now, sometimes that written opportunity is pretty lacking. So, it helped me focus more on a regular basis and I liked that part of it.

Efficiency

Another concept prevalent in interviews was the efficiency of NC Write. One efficiency that teachers cited as beneficial to students' growth in writing quality was the affordance of writing practice opportunities. An NC + TRAD teacher explained, "The more you write, the better you're going to be, so it just gives them the opportunities that they need." Teachers in both conditions indicated that students were able complete more writing with the program than they were otherwise. One dimension of this efficiency was that all text in NC Write was typed, which allowed for faster production of text than writing longhand. One NC + TRAD student explained, "When I'm on the computer, I have more time because I can type faster than writing." An NC + SRSD attributed the additional writing practice to increased comfort with writing production processes, particularly when applied across content areas.

Teachers also described how automated feedback was efficient in saving them time. An NC + TRAD teacher indicated, "I think that's the number one selling point for me: that I didn't have to sit there and read every single [essay] to find all the grammar mistakes, the spelling mistakes, that kind of thing." The lessons, prompts, graphic organizers, and portfolios associated with the program provided additional efficiencies. One NC + TRAD teacher shared, "I liked that I didn't have to spend an hour planning an essay for them to write and lessons on sentence structure and that it was already there for me." Teachers described efficiencies over other classroom practices and their own grading and provision of feedback. One NC + TRAD teacher described how she used NC Write as a formative assessment component of her writing instruction:

> I like to model and if we write a paper together . . . that doesn't show me where each individual student really is with their writing. So when we do a few practices together, and then they write it on their own on the computer, it really lets me know how much they've learned in that past week. This makes them accountable for their own writing.

NC Write afforded a range of efficiencies, including additional writing practice opportunities for students and time-saving benefits for teachers.

Revealing evidence of growth

When asked whether NC Write helped them to become better writers, students in both conditions spoke to the fact the program provided them with visible evidence of growth. Students were generally cognizant about the quality of their writing and the extent to which it improved during the time they used NC Write. For example, one NC + TRAD student explained that his writing quality became "a lot better. Because when I first did it, I made a 12, and I went to a 23. My lowest one in the bar graph is my Sentence Structure, like my commas and how strong my sentences are." An NC + SRSD student provided a similar perspective, explaining:

> At the beginning, my score was 18. And with feedback and how it evaluates—
> it tells you what you need to do, what things to help you—I've learned from
> that, and now my top score is 24.

Students described using their portfolios to monitor growth, and liked that they had objective measures of their writing quality and growth. One NC + TRAD student shared:

> Being able to see my growth instead of people saying 'oh yeah you did good'
> . . . you could actually see yourself climbing in your writing instead of people
> having to tell you.

Further, interview results provided some evidence that the availability of visible evidence of growth emphasized the relations among essay elements and writing quality to students. For example, the automated scores and feedback made explicit to students that essays with more essay elements, and/or fewer errors, tended to be of a higher quality. One NC + TRAD student described his experience:

> When it said choose more upper grade words or make your sentence longer
> or do some more paragraphs . . . [I] revised it, and then I put more stuff—
> what I needed—and then when I hit 'submit' again, I got a higher score.

Writing quality feedback

The most common concept addressed in interviews related to NC Write's automated feedback. Nearly all students described first examining the PEG total essay score, then the trait scores, each time they received feedback. However, beyond these feedback components, students used disparate strategies to approach and apply the feedback. Some students reviewed only the total essay and trait scores, ignoring the other feedback components. At times, these students used the trait scores to focus their improvement efforts. For example, one NC + TRAD student described how he used the accompanying bar graph to compare trait scores. "I look at the lowest one, and I try to improve that to at least the next line— say if it was at a 3 . . . I'd try to go for a 4 next."

Most students reviewed and applied the spelling and grammar feedback after examining the total essay and trait scores. An NC + SRSD student described this process as follows:

> I look at the things that I spelled wrong and I go back and think of how I
> spelled it wrong . . . and then I would get the grammar parts and it would be
> saying that my sentences [were] too long and then I would just review the
> whole thing and sometimes I would revise it.

In general, students found the spelling and grammar feedback accessible and actionable as it related to specific text in their essays.

Students less frequently reported using the writing analysis evaluation and feedback. This appeared to be due to the density and complexity of this feedback, which included numerous suggestions for each trait. Students who reviewed this feedback varied in the extent to which they engaged with it. For example, one NC + TRAD student shared, "The text, I would skim over it, but I wouldn't read in depth."

Most students who used the writing analysis seemed to consider the feedback generally rather than examine or apply it in a systematic way. An NC + SRSD student explained, "Though you might not take notes on it, you'll just look at it and be like 'OK, keep this in mind so next time I write it, I know that I need to add more details or more supporting evidence.'" Students in both conditions also noted that since all feedback was recorded in their portfolios they could reference previous feedback as needed.

In their interviews, teachers in both conditions confirmed that students focused on the total essay scores and underutilized the writing analysis evaluation and feedback. However, teachers shared that students' score comparisons eventually led students to compare the content of their writing and consider the relations between text features and scores. Teachers found it necessary to redirect students to the additional feedback available, but did not dissuade students from comparing total essay scores.

A second property of the writing quality feedback was that teachers had to provide significant support in order for students to effectively interpret and apply the feedback. For example, one NC + TRAD teacher indicated, "If you don't sit down with your students and explain [the feedback] to them . . . I don't think they can understand it." A number of students confirmed that they needed help understanding the feedback. One NC + SRSD student recalled, "At first I didn't know what half that stuff was." The need for support was partly attributed to students' limited vocabulary knowledge. In their interviews, teachers in both conditions characterized the provision of this support as part of their instructional responsibilities and not a particular deficiency of the feedback. An NC + TRAD teacher explained, "The complexity of the wording was difficult for them, but it's necessary." Teachers described using both content and task scaffolding to support students' interpretation and application of feedback and indicated that struggling writers required the most support.

Though students needed considerable support to apply the feedback, it enabled teachers to shift the nature of their support from evaluation to coaching. One NC + TRAD teacher explained, "The feedback sometimes offered me an opportunity to be one-on-one with [students] and say, 'Well, now this is why this is highlighted,' or 'This is why they wanted you to do this,' and things of that nature." Provided by an external agent, the feedback provoked communication and collaboration among students and teachers regardless of instructional context.

Finally, prevalent writing challenges explicated by feedback solicited teacher support of an instructional nature. In their interviews, teachers described the importance of regularly reviewing the automated feedback provided to students to inform instructional decisions. An NC + TRAD teacher explained:

> In the feedback for most of your students, it's going to be the same because most of them have the same weaknesses and then you'll see it all throughout every single class. The fact that they're all getting the same feedback . . . it allows you to take a lesson and say, "OK, here's the feedback that everyone's getting, so let's see what we can do to make that better.

In sum, though feedback required considerable support from teachers, it enabled teachers to adopt more of a coaching role and allowed them to identify instructional priorities.

A third property of writing quality feedback was that there were specific ways in which it supported improvements in students' writing performance. In their interviews, students reported adding more details to their essays, using greater specificity in their writing, improving spelling/grammar, word choice, and sentence structure, and improving self-monitoring and self-evaluation of their writing due to the feedback they received. Students attributed the quality and quantity of feedback to improvements in their writing performance. One NC + SRSD student explained, "It's just more feedback than an actual teacher would give you, I believe." Another NC + SRSD student felt, "It's more detailed and it explains it a lot better than the teachers can sometimes." In their interviews, some students described that the specificity of feedback allowed them to improve their writing. One NC + SRSD student explained, "It makes it very obvious what you did or didn't do right."

Some teachers also noted that receiving regular and immediate feedback for each essay draft appeared to support students' self-monitoring. One NC + TRAD teacher described how, over time, her students became better at correcting their own errors as a result of receiving regular feedback. She shared:

> Just being able to go back and do some things . . . before I tell them to edit, they are looking for [errors] or see them . . . . I feel like my kids started to see that better and were able to make those adjustments, and they started to have less errors the more we did it, and I think that was effective.

Limitations of the framework

NC Write's affordances in two instructional contexts—i.e., supporting writing instruction, efficiency, writing quality feedback, and evidence of growth—appeared to support instructional environments that emphasized this framework of deliberate writing practice. However, interview results also revealed that there were aspects of NC Write that limited the effectiveness of the same framework. These aspects related to limitations of NC Write's automated feedback and interactive lessons, as well as its inability to identify plagiarized writing.

There were a number of ways in which the writing quality feedback diverged from the principles of effective formative feedback (e.g., Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Parr & Timperley, 2010; Shute, 2008). This divergence limited the relevance and efficacy of the feedback. For instance, some students described the writing analysis evaluation and feedback as too long and complex. One NC + TRAD student conceded, "Some of the feedback I didn't quite understand when they used complicated words." This concern was echoed by teachers in both conditions, a number of whom described the feedback as differentially effective. For example, one NC + SRSD teacher shared:

> My kids who were at a higher level, it probably helped them more because they could make more sense of it. My kids that read at a lower level probably

had some difficulty understanding exactly what it was trying to get across to them as far as what they did well and what they should work on.

Another divergence from the principles of effective formative feedback related to the feedback specificity. Some students indicated the writing analysis was too vague. One NC + TRAD student recommended, "One thing I would change is how it would be a little more specific about how I could change things . . . it could point out what I should change and how I can make things better." Some students similarly observed a lack of variety in aspects of the writing analysis. These opinions were shared by teachers, who found the feedback accurate and generally useful though insufficiently specific to support continuous improvements from draft to draft. Consistent with the quantitative findings, teachers and students in both conditions described a point where students reached a writing quality ceiling and struggled to improve further. An NC + TRAD student described, "I made a 24 on the last one . . . . I've even rewritten an entire paragraph to see if that would help or anything and I feel like I cannot reach that point above a 24." One NC + SRSD teacher suggested restructuring the format of the feedback to be more oriented to the revision process. She explained, "It's almost as if it needs to be laid out in a way that it's speaking to the students: 'OK, here's your score. What does that mean? What do I do now?'"

Finally, there was some evidence the presentation of the feedback diverged from the principles of effective formative feedback, limiting feedback efficacy. Due to the visual nature of the spelling/grammar feedback, some students corrected the spelling/grammar errors via surface-level edits, believing they had made sufficient revisions to significantly improve the quality of their essays. Consequently, students in both conditions described making edits and being discouraged by the lack of change in scores or subsequent feedback. In contrast, the writing analysis was less visual, allowed more subjective interpretation, and it was with the traits of development of ideas, organization, and style that students appeared to have the greatest difficulty enacting feedback. One NC + SRSD teacher explained:

> One of my kids, she would get a little upset because they were like, "Well, I went through and I did what they asked me to do and I still got an 18." Once they went back and looked at the feedback again . . . . "You're kind of still making the same mistake that you were making before. So, you're looking at the feedback, but are you really applying the feedback or just changing something and thinking 'Okay, that's going to make it better?'"

The writing quality feedback diverged from the principles of effective formative feedback in terms of complexity, specificity, and presentation (Shute, 2008) which limited its utility and efficacy.

Teachers and students shared mixed opinions regarding the lessons in NC Write during interviews. Some students recalled particular lessons they attributed to helping them improve as writers. Students generally liked the interactive nature of the lessons. Some teachers also liked the lessons and found them effective. One NC + SRSD teacher shared, "The mini-lessons . . . helped out tremendously. They were not too long, they were straight to the point, giving the information, and it was in a kid-friendly way where they could understand it, whatever the topic was." However,

the lessons solicited greater criticism from students and teachers than any other component of NC Write. The most common concerns raised by students were that the lessons should have been less challenging and less dense with information. Teachers' primary criticism of the lessons was that they did not provide sufficient data about student performance. For example, one NC + TRAD teacher explained:

> I would like a little more feedback on how the kids are doing on the lessons as they work through them and when they have actually completed a lesson."

In their interviews, students and teachers were asked what they would change about NC Write. Many of students' recommendations addressed existing features of NC Write that were not utilized for the present study, such as peer review functionality. Both students and teachers suggested that additional feedback in the form of example essays would have been helpful. NC Write included some annotated example essays at the time of the intervention but these were limited in number and not clearly linked to total essay scores. Many teachers indicated that NC Write needed to be able to identify text plagiarized from other writers (i.e., copied from essays found on the internet), or text copied from the prompts and source materials. Plagiarism was a greater concern in the upper grades, though sixth grade teachers reported instances of students copying the prompts. One NC + SRSD teacher explained:

> All they did was copy the prompt and paste it and they got a 12. So, that was, of course, a negative because they didn't even write anything on their own.

Though students had been informed that accurate PEG scoring was dependent on "good faith" essays, plagiarism was a frustration for teachers as unobserved cases of plagiarism resulted in students having reduced writing practice opportunities and receiving potentially misleading feedback about their writing performance.

In sum, limitations of the framework for deliberate writing practice included the extent to which feedback diverged from the principles of effective feedback, limited lesson data, and the program's lack of a plagiarism scanner, all of which limited its effectiveness in process writing and strategy instructional contexts.

3.6 Discussion

In this study, we adopted a mixed-methods design to examine the implementation and integration of AWE with two different approaches to teacher-led writing instruction at the middle-school level. One instructional context integrated AWE within a traditional process approach to writing instruction, while the other context integrated AWE with strategy instruction. We evaluated both instructional contexts with respect to fostering growth in students' first-draft writing quality across successive essays over time. Further, we examined students' and teachers' experiences with and perceptions of the AWE system to help explain the quantitative findings, and further understand benefits and limitations of AWE as perceived by users, as well as examine what these perceptions indicated regarding

the affordances of AWE in the different instructional contexts. We discuss both research questions in turn.

What are students' first-draft writing performance (i.e., writing quality, essay length, and essay elements) growth trajectories when AWE is used within two instructional contexts: process writing instruction and strategy instruction?

Prior research has rarely examined growth in first-draft writing performance across successive essays and found evidence of transfer. The present study extends existing AWE research by documenting, across multiple writing performance outcomes, that students in different AWE-supported instructional contexts improved their writing quality, the length of their essays, and the number of basic elements they included in their first-drafts. Importantly, results indicated that students in both instructional contexts exhibited comparable rates of growth in writing performance: both NC + TRAD and NC + SRSD students exhibited a period of accelerated growth in writing performance that slowed over time, plateauing after the fourth first-draft essay written during the intervention. Though the present study focused on *teacher-directed* instructional formats, findings are consistent with Roscoe and colleagues' work studying the effects of different *student-directed* practice formats for utilizing AWE: students in traditional process writing practice formats, strategy-based practice, and game-based practice equally improved the quality of their essays when revising with AWE (Roscoe et al., 2013, 2018, 2019).

Though strategy instruction is associated with large effect sizes (Graham et al., 2012; Graham & Perin, 2007), and SRSD-based interventions have produced, on average, the largest effect sizes for writing quality of all writing interventions (Graham & Harris, 2018), there are several possible explanations for why the two groups of students, using AWE in different instruction contexts, exhibited comparable growth trajectories. The first is that adapting SRSD to a lower-intensity format to support teacher implementation and reduce professional development constrained the net efficacy of the approach. However, Palermo and Thomson (2018) found a large, positive effect at posttest of NC + SRSD on students' writing quality (Cohen's d_z = 1.18), essay length, (d_z = 1.36), and essay elements (d_z = .97), comparable to effect sizes observed in other SRSD studies (e.g., Graham & Harris, 2003; Graham, 2006; Graham & Perin, 2007; Graham et al., 2012). A second explanation is that results are partially attributable to between-condition mean differences in AWE exposure. However, all final conditional models included a student-level control for AWE exposure, which adjusted for both between- and within-condition differences in exposure to NC Write. A third explanation relates to the frequency of measurement. As Palermo and Thomson (2018) found that NC + SRSD students demonstrated greater gains in writing performance than NC + TRAD students from pretest to posttest, in the present study it may be that mean changes in writing performance across first-draft essays were too minor—and/or standard errors too large—to identify between-condition differences. A final explanation, supported by the interview results, is that NC Write offered homogeneous affordances (and limitations) across instructional contexts, moderating the effect of writing instruction on writing performance. This final hypothesis is interrogated further below.

What are students' and teachers' experiences with and perceptions of the AWE system and what do these perceptions indicate regarding the affordances of AWE in the different instructional contexts?

Though NC Write was implemented within process writing instruction and strategy instruction contexts, students and teachers described similar experiences with and perceptions of NC Write across the two instructional contexts. This finding suggested that NC Write provided homogeneous affordances and limitations in the different instructional contexts.

Qualitative data analysis revealed that, regardless of instructional context, NC Write provided a framework for deliberate writing practice. In this framework, students' growth in writing performance was supported by a cycle of learning, practice, and feedback. Specifically, NC Write enabled deliberate practice by supporting teachers' writing instruction, offering efficiencies not otherwise possible, affording writing quality feedback, revealing evidence of growth, and motivating continued and iterative engagement with cycles of practice and feedback.

The SRSD model was identified as an ideal form of strategy instruction to implement in the present study given that SRSD is an evidence-based practice (Harris & Graham, 2016) and a particularly effective form of strategy instruction as it includes self-regulation instruction (Graham at al., 2012). While process writing instruction shared some similarities to SRSD instruction, it differed in the extent to which it emphasized strategies for planning, writing, and revising argumentative essays; included the teaching of self-regulation procedures; and was criterion based and/or instructionally responsive. Regardless, interview results suggested that the negligible differences observed between process writing and strategy instruction on students' writing performance could be explained by the affordances of NC Write that were common to both conditions.

For example, central to SRSD instruction is the teaching of planning and writing strategies (Graham & Harris, 2018). The purpose of the **DARE** strategy, described previously, is to help students recall and incorporate all basic elements when composing argumentative essays. This is achieved by instructing students to compare their essays to a list of criteria for high-quality argumentative writing. Interestingly, though process writing instruction included less of an emphasis on planning, writing, and revising strategies, interview results suggested that AWE led students—in both conditions—to examine the relations among essay elements and the writing quality scores assigned by PEG. Thus, the practice of comparing essays and scores supported students' understanding of the relations between text features and writing quality, fostering a better understanding of the components of good argumentative writing.

Another hallmark of SRSD instruction is the teaching of self-regulation procedures and positive beliefs about writing to help students apply writing strategies, manage the writing process, and collect visible evidence of their growth (Harris, Graham, & Mason, 2006). Though NC + TRAD students were not explicitly taught self-regulation procedures and positive beliefs about writing, interview results provided moderate evidence that AWE provided students in both conditions with tools to help them manage their use of writing strategies and the writing task

and allowed students to collect evidence of their writing growth. When the opportunity to practice repetitively and receive feedback allows for deliberate practice, students can improve their ability to self-monitor, control, and self-evaluate their performance (Ericsson, 2006). In the context of AWE, the cycles of practice and feedback were seen to support students' self-monitoring and self-evaluation. When students learned a new instructional strategy (from a lesson in NC Write or from their classroom teacher), they were able to promptly try the instructional strategy and observe the impact on their writing quality. That automated scores and feedback were immediately available and generally perceived as valid and valuable may have further fostered students' self-evaluation. Automated feedback provided all students with concrete evidence of growth; this was highly visible to students on a draft to draft basis and manageable for students to track over time via their writing portfolios. This evidence of growth, along with mastery experiences over cycles of practice and feedback, provided students with a source of efficacy information. Thus, AWE appeared to provide some of the same levers as SRSD instruction related to the development of self-regulatory skills and abilities and positive beliefs about writing.

Finally, SRSD instruction differs fundamentally from process writing instruction in the extent to which it is criterion-based and instructionally responsive. Though process writing instruction tends not to be as tailored to the needs of students as SRSD instruction, interview results suggested the interconnected nature of the learning, practice, and feedback cycles afforded by AWE explicated both current performance and the components necessary for improved performance for all students. This information allowed students to concentrate on improving specific aspects of their writing performance. Collectively, the cycles of learning, practice, and feedback may have helped demystify the construct of writing quality for students.

Thus, results are promising with respect to using AWE to support the teaching and learning of writing. The comparable growth trajectories exhibited by students in the two treatment conditions suggests that AWE can be implemented effectively in a variety of instructional contexts. When so implemented, students appear to improve their writing performance and teachers and students identify common affordances of AWE. In sum, our results suggest effective writing instruction with AWE looks markedly different than typical process approaches to writing *or* strategy instruction; that is, the use of AWE appears to transform instruction in ways that align with a framework for deliberate writing practice.

Limitations of the Present Study

Several limitations should be considered when interpreting study results. Students were assigned to condition by teacher, rather than randomly. Consequences of this quasi-experimental design included between-condition initial differences in writing quality as well as between-condition differences in the proportion of Black students and SWDs. Resource limitations precluded recording or observing teachers' writing instruction to confirm NC + TRAD teachers' descriptions of their process writing instruction. Additionally, AWE exposure was limited to a measure of activities completed as captured by NC Write log-files. Lesson activity logs, for example,

tracked which lessons students accessed and for how long, but did not include measures of students' performance during lessons. While participant selection criteria included lack of exposure to NC Write, interview results suggested students and teachers required some time to develop familiarity with the software which may have impacted outcomes. Students' writing performance may have been depressed toward the end of the intervention and at posttest in part due to the timing of the intervention—toward the end of the school year—when student effort tends to wane. We are not able to disentangle the effects of timing from those inherent to the intervention or NC Write. Future research should examine whether results are similar if treatment occurs earlier in the school year as well as examine maintenance effects.

4. Indications for Future Developments

Our findings showed that though AWE's benefits were not context dependent, AWE's limitations were also insensitive to context, as NC + TRAD and NC + SRSD teachers and students reported similar challenges and frustrations. This finding suggests that teachers were not able to mitigate the limitations of AWE by varying instruction. Thus, it is important for teachers to recognize that using AWE may result in their needing to expend more effort to help students work with and around the limitations of AWE. This in turn will require effective professional development and support that goes beyond the technical knowledge required to use AWE.

Results of the present study therefore support several indications for future developments. Future research is needed to understand how teachers should change and adapt writing instruction when it includes AWE. As the focus of the extant AWE research is those components of instruction that teachers can reduce or omit with AWE (i.e., the time-saving aspects of AWE), future development should focus on components of instruction that become increasingly demanding with AWE. For example, in the context of AWE, teachers likely need to spend more time teaching students the qualities of good writing, providing students with feedback in the form of examples and counterexamples, and helping students connect general feedback to specific actions they can take to improve—in short, providing more evidence-based writing instruction (e.g., Graham & Perin, 2007). Additionally, professional development related to AWE should stress the importance of changing and adapting instruction when AWE is part of teachers' writing toolkits (see also Knight et al., 2020).

In the present study we analyzed first draft scores for each essay, and results indicated that students using AWE within process writing instruction and strategy instruction contexts improved their writing performance at comparable rates. However, it is possible that there were differences in the ways that students in the two conditions subsequently revised their essays as well as the way that students in each condition enacted their revision process. Additional research should examine writing performance associated with both first drafts and final drafts to better understand potential differences attributable to instructional contexts. A promising method of doing so is to leverage feedback available from keystroke logging software that reveals evidence of students' drafting and revision processes (see Vandermeulen, Leijten, & Van Waes, 2020).

Finally, to further encourage deliberate writing practice, students require carefully designed practice opportunities that can be mastered sequentially (Ericsson, 2006). In the context of AWE, such practice opportunities should allow students to apply specific skills and strategies in their writing, but not require production of an essay to receive feedback. For example, one practice activity might focus specifically on developing a claim and another on refuting counter-claims (or producing purpose-specific introductions, elaboration, and/or conclusions). Collectively, such activities should better support students' maintenance of writing performance growth over time. Attention should be given to the development of these practice activities and the associated AES models, and future research should examine the role of such practice opportunities in the context of writing instruction with AWE.

Acknowledgements

References

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Erlbaum.

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. In E. Butterfield (Ed.), *Children's writing: Toward a process theory of the development of skilled writing* (pp. 57–81). Greenwich, CT: JAI Press.

Brindle, M., Graham, S., Harris, K. R., & Hebert, M. (2016). Third and fourth grade teacher's classroom practices in writing: A national survey. *Reading and Writing, 29*(5), 929–954. https://doi.org/10.1007/s11145-015-9604-x

Charlton, C., Rasbash, J., Browne, W. J., Healy, M. and Cameron, B. (2017) *MLwiN Version 3.02.* Centre for Multilevel Modelling, University of Bristol.

Clare, L., Valdés, R., & Patthey-Chavez, G. G. (2000). Learning to write in urban elementary and middle schools: An investigation of teachers' written feedback on student compositions (Center for the Study of Evaluation Technical Report No. 526). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing (CRESST). https://doi.org/10.4135/9781412950558.n70

Clark, V. L. P., & Creswell, J. W. (2008). *The mixed methods reader.* Thousand Oaks, CA: Sage Publications.

Corbin, J. & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). Thousand Oaks, CA: Sage. https://doi.org/ 10.4135/9781452230153

Cotos, E., Huffman, S., & Link, S. (2020). Understanding graduate writers' interaction with and impact of the Research Writing Tutor during revision. *Journal of Writing Research 12*(1), 187-232. https://doi.org/10.17239/jowr-2020.12.01.07

Creswell, J. (2015). *A concise introduction to mixed methods research.* Thousand Oaks, CA: Sage Publications.

Dikli, S. (2010). The nature of automated essay scoring feedback. *CALICO Journal, 28*(1), 99–134. https://doi.org/10.11139/cj.28.1.99-134

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J.

Feltovich (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 39–68). New York: Cambridge University Press. https://doi.org/10.1017/cbo9780511816796.038

Festas, I., Oliveira, A. L., Rebelo, J. A., Damião, M. H., Harris, K., & Graham, S. (2015). Professional development in self-regulated strategy development: Effects on the writing performance of eighth grade Portuguese students. *Contemporary Educational Psychology*, *40*, 17–27. https://doi.org/10.1016/j.cedpsych.2014.05.004

Flower, L., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. Gregg and E. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31–50). Hillsdale, NJ: Erlbaum

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. https://doi.org/10.2190/DH8F-QJWM-J457-FQVB

Gilbert, J., & Graham, S. (2010). Teaching writing to elementary students in Grades 4–6: A national survey. *Elementary School Journal*, *110*, 494–518. https://doi.org/10.1086/651193

Graham, S. (2006). Strategy instruction and the teaching of writing: A meta-analysis. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 187–207). New York: Guilford. https://doi.org/10.1007/s11145-008-9121-2

Graham, S. (2018). A writer(s) within community model of writing. In C. Bazerman, V. Berninger, D. Brandt, S. Graham, J. Langer, S. Murphy, P. Matsuda, D. Rowe, & M. Schleppegrell,(Eds.), *The lifespan development of writing* (pp. 272-325). Urbana, IL: National Council of English.

Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323–344). New York: Guilford Press.

Graham, S., & Harris, K. R. (2018). An examination of the design principles underlying a Self-Regulated Strategy Development study. *Journal of Writing Research*, *10*(2), 139–187. https://doi.org/10.17239/jowr-2018.10.02.02

Graham, S., Harris, K. R., Fink-Chorzempa, B., & MacArthur, C. (2003). Primary grade teachers' instructional adaptations for struggling writers: A national survey. *Journal of Educational Psychology, 95*(2), 279–292. https://doi.org/10.1037/0022-0663.95.2.279

Graham, S., Harris, K. R., Fishman, E., Houston, J., Wijekumar, K., Lei, P. W., & Ray, A. B. (2019). Writing skills, knowledge, motivation, and strategic behavior predict students' persuasive writing performance in the context of robust writing instruction. *The Elementary School Journal, 119*(3), 487-510. https://doi.org/10.1086/701720

Graham, S., Harris, K. R., & McKeown, D. (2013). The writing of students with LD and a meta-analysis of SRSD writing intervention studies: Redux. In L. Swanson, K. R. Harris, & S. Graham (Eds.). *Handbook of learning disabilities* (pp. 405–438). (2nd ed.). NY: Guilford Press. https://doi.org/10.1111/ldrp.12004

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115*, 523–547. https://doi.org/10.1086/681947

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*(4), 879–896. https://doi.org/10.1037/a0029185

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*, 445–476. https://doi.org/10.1037/0022-0663.99.3.445

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), 4–43.

Harris, K., & Graham, S. (1988). Self-instructional strategy training: Improving writing skills among educationally handicapped students. *Teaching Exceptional Children*, *20*, 35–37. https://doi.org/10.1177/004005998802000211

Harris, K. R., & Graham, S. (2016). Self-regulated strategy development in writing: Policy implications of an evidence-based practice. *Policy Insights from the Behavioral and Brain Sciences*, *3*, 77–84. https://doi.org/10.1177/2372732215624216

Harris, K. R., Graham, S., Brindle, M., & Sandmel, K. (2009). Metacognition and students' writing. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 131-153). Mahwah, NJ: Lawrence Erlbaum.

Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with

and without peer support. *American Educational Research Journal, 43*(2), 295–340. https://doi.org/10.3102/00028312043002295

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112. https://doi.org/10.3102/003465430298487

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.) *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Mahwah, NJ: Erlbaum.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of Writing Research, 1*, 1–26. https://doi.org/10.1080/00461520903213600

Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist, 44*(4), 250–266. https://doi.org/10.1080/00461520903213600

Kellogg, R. T., Whiteford, A. P., & Quinlan, T. (2010). Does automated feedback help students to write? *Journal of Educational Computing Research, 42*, 173–196. https://doi.org/10.2190/EC.42.2.c

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Knight, S., Shibani, A., Abel, S., Gibson, Ryan, P., Sutton, N., Wight, R., Lucas, S., Sándor, Á., Kitto, K., Liu, M., Vijay Mogarkar, R., & Buckingham-Shum, S.J. (2020). AcaWriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research 12*(1), 141-186. https://doi.org/10.17239/jowr-2020.12.01.06

Lepper, M. R., & Hodell, M. (1989). Intrinsic motivation in the classroom. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 73–105). San Diego, CA: Academic Press.

Matsumura, L. C., Patthey-Chavez, G. G., Valdés, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower-and higher-achieving urban schools. *The Elementary School Journal, 103*, 3–25. https://doi.org/10.1086/499713

Mayfield, E., Adamson, D., Woods, B., Miel, S., Butler, S., & Crivelli, J. (2018). Beyond automated essay scoring: Forecasting and improving outcomes in middle and high school writing. *Proceedings of the 8th International Conference on Learning Analytics & Knowledge,* 1–8. Retrieved from https://help.turnitin.com/Resources/RA%20Curriculum%20Resources/Research/Revision%20Assistant%20Efficacy%20LAK%202018.pdf

McCutchen, D. (1988). "Functional automaticity" in children's writing: A problem of metacognitive control. *Written Communication, 5*(3), 306–324. https://doi.org/10.1177/0741088388005003003

McKeown, D., Brindle, M., Harris, K. R., Graham, S., Collins, A. A., & Brown, M. (2016). Illuminating growth and struggles using mixed methods: Practice-based professional development and coaching for differentiating SRSD instruction in writing. *Reading and Writing, 29*(6), 1105–1140. https://doi.org/10.1007/s11145-016-9627-y

McKeown, D., FitzPatrick, E., Brown, M., Brindle, M., Owens, J., & Hendrick, R. (2018). Urban teachers' implementation of SRSD for persuasive writing following practice-based professional development: Positive effects mediated by compromised fidelity. *Reading and Writing,* 1–24. https://doi.org/10.1007/s11145-018-9864-3

Morgan, J., Shermis, M. D., Van Deventer, L., & Vander Ark, T. (2013) Automated Student Assessment Prize: Phase 1 & Phase 2. Retrieved from http://gettingsmart.com/wpcontent/uploads/2013/02/ASAP-Case-Study-FINAL.pdf

Morphy, P., & Graham, S. (2012). Word processing programs and weaker writers/readers: A meta-analysis of research findings. *Reading and Writing, 25*(3), 641–678. https://doi.org/10.1007/s11145-010-9292-5

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*(4), 375–401. https://doi.org/10.1007/s11251-008-9053-x

Palermo, C., & Thomson, M. M. (2018). Teacher implementation of Self-Regulated Strategy Development with an automated writing evaluation system: Effects on the argumentative

writing performance of middle school students. *Contemporary Educational Psychology, 54,* 255–270. https://doi.org/10.1016/j.cedpsych.2018.07.002

Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing, 15,* 68–85. https://doi.org/10.1016/j.asw.2010.05.004

Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology, 108*(8), 1098–1120. https://doi.org/10.1037/edu0000103

Patthey-Chavez, G. G., Matsumura, L. C., & Valdés, R. (2004). Investigating the process approach to writing instruction in urban middle schools. *Journal of Adolescent & Adult Literacy,* 462–476.

Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? *Computer Assisted Language Learning, 31,* 653–674. https://doi.org/10.1080/09588221.2018.1428994

Roscoe, R. D., Allen, L. K., Johnson, A. C., & McNamara, D. S. (2018). Automated writing instruction and feedback: Instructional mode, attitudes, and revising. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 2089–2093. https://doi.org/10.1177/1541931218621471

Roscoe, R. D., Allen, L. K., & McNamara, D. S. (2019). Contrasting writing practice formats in a writing strategy tutoring system. *Journal of Educational Computing Research, 57,* 723–754. https://doi.org/10.1177/0735633118763429

Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A., & McNamara, D. S. (2014). The Writing Pal intelligent tutoring system: Usability testing and development. *Computers & Composition, 34,* 39–59. https://doi.org/10.1016/j.compcom.2014.09.002

Roscoe, R. D., & McNamara, D. S. (2013). Writing Pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology, 105,* 1010–1025. https://doi.org/10.1037/a0032340

Roscoe R. D., Snow E. L., McNamara D. S. (2013). Feedback and revising in an intelligent tutoring system for writing strategies. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.). *Artificial Intelligence in Education. AIED 2013. Lecture Notes in Computer Science, vol 7926.* Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39112-5_27

Scardamalia, M., Bereiter, C., & Goleman, H. (1982). The role of production factors in writing ability. In M. Nystrand (Ed.), *What writers know: The language, process, and structure of written discourse* (pp. 175–210). San Diego, CA: Academic Press. https://doi.org/10.1017/s0047404500010095

Shermis, M. D., Garvan, C. W., & 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.

Shermis, M. D., & Hamner, B. (2013). Contrasting state-of-the-art automated scoring of essays. In M. D. Shermis, & J. Burstein (Eds.), Handbook of automated essay evaluation: Current applications and new directions (pp. 313–346). New York, NY: Routledge. https://doi.org/10.4324/9780203122761.ch19

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153–189. https://doi.org/10.3102/0034654307313795

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Thousand Oaks, CA: Sage. https://doi.org/10.4135/9781452230153

Stevenson, M. (2016). A critical interpretative synthesis: The integration of automated writing evaluation into classroom writing instruction. *Computers and Composition, 42,* 1–16. https://doi.org/10.1016/j.compcom.2016.05.001

Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19,* 51–65. https://doi.org/10.1016/j.asw.2013.11.007

Vandermeulen, N., Leijten, M., & Van Waes, L. (2020). Reporting writing process feedback in the classroom: Using keystroke logging data to reflect on writing processes. *Journal of Writing Research 12*(1), 109-140. https://doi.org/10.17239/jowr-2020.12.01.05

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction, 22,* 333–362. https://doi.org/10.1207/s1532690xci2203_3

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal, 3,* 22–36. https://doi.org/10.1080/15544800701771580

Wilson, J. (2017). Associated effects of automated essay evaluation software on growth in writing quality for students with and without disabilities. *Reading and Writing, 30,* 691-718. https://doi.org/10.1007/s11145-016-9695-z

Wilson, J., & Czik, A. (2016). Automated essay evaluation software in English language arts classrooms: Effects on teacher feedback, student motivation, and writing quality. *Computers and Education, 100,* 94-109. https://doi.org/10.1016/j.compedu.2016.05.004

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

Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research, 58,* 87-125. https://doi.org/10.1177/0735633119830764