ABSTRACT


Process-based approaches to writing tend to overlook the self-regulatory skills and motivational beliefs required for proficient writing (Harris, Santangelo, & Graham, 2008) and do not provide the support many students need to develop into effective writers (Graham, Harris, & Mason, 2005; Harris, Graham, & Mason, 2006). Additionally, restricted writing opportunities preclude the sustained deliberate practice students need to develop expertise in writing (Kellogg & Whiteford, 2009). This study examined an intervention that incorporated the self-regulated strategy development model (SRSD, Graham & Harris, 1993) with the automated writing evaluation (AWE) program NC Write. An embedded quasi-experimental mixed methods design was used to determine the impact of the intervention on students’ argumentative writing performance, knowledge, and self-efficacy. Middle school students \((N=829)\) participated in one of three conditions: NC Write + traditional writing instruction, NC Write + SRSD instruction, or a comparison condition.

Results of multi-level models that controlled for pretest performance and predicted posttest performance averaging across students and within teachers showed that students in the NC Write + SRSD instruction condition produced posttest essays that were of a higher quality, longer, and included more basic elements of argumentative essays than students in the other two conditions. Students in the NC Write + traditional writing instruction condition produced higher-quality essays than students in the comparison condition at posttest. Students in the NC Write + SRSD instruction condition identified more essay elements at
posttest, though there were no between-condition differences in writing knowledge of substantive processes or in students’ writing self-efficacy at posttest.

Additional multi-level models were specified to include all essays written by treatment condition students and examine the shape of growth in writing performance. Results showed that students’ growth in writing quality, essay length, and essay elements was best represented by a quadratic growth model. On average, students’ growth in writing performance reached a plateau following the fourth essay written during the intervention. Differences in rates of change and deceleration in writing quality, essay length, and essay elements were not significantly different between the two treatment conditions.

Survey results showed students and teachers held generally favorable opinions of NC Write. Interview results determined that NC Write as well as the overall writing intervention had acceptable social validity. Qualitative data analysis revealed that NC Write provided a framework for deliberate writing practice. In this framework students’ growth in writing performance is supported by a cycle of learning, practice, and feedback. NC Write enabled deliberate practice by affording writing quality feedback, efficiency, and evidence of growth, and supporting teachers’ writing instruction and students’ intrinsic motivation. Limitations of the framework included some aspects of feedback, limited lesson data, and lack of a plagiarism scanner in NC Write. Implications from these findings support integrating SRSD instruction with an AWE program to support teacher implementation of the SRSD model and more efficiently provide students with the strategy instruction, practice opportunities, and feedback needed to develop proficiency in writing. Recommendations are provided for AWE programs to better support students’ maintenance of writing quality growth.
A Framework for Deliberate Practice: Self-Regulated Strategy Development and an Automated Writing Evaluation Program

by

Corey John Palermo

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Curriculum & Instruction

Raleigh, North Carolina

2017

APPROVED BY:

Dr. Margareta Thomson

Committee Chair

Dr. John L. Nietfeld

Dr. Jessica DeCuir-Gunby

Dr. Andrew McEachin

Dr. Meg Blanchard
BIOGRAPHY

Corey Palermo began his educational career as a lateral entry teacher in the North Carolina public school system. He completed his North Carolina teaching certification coursework at North Carolina Central University in 2006. Mr. Palermo’s interest in educational assessment led him to accept a position with Measurement Incorporated in 2007 where he helped to develop high-stakes assessments for a number of state departments of education. Mr. Palermo earned a M.Ed. in Curriculum and Instruction from North Carolina State University in 2010 and was subsequently accepted into the Educational Psychology Ph.D. program. He continues to work for Measurement Incorporated as Vice President, Performance Assessment Scoring, where he directs a team of program managers to provide management and operations oversight for multi-state assessment programs administered in 15 states and U.S. territories.
ACKNOWLEDGEMENTS

To my wife, Ashley, for years and years of patience.

Also to my family for the support and encouragement.

Thanks to Dr. Margareta Thomson and Dr. John L. Nietfeld for their mentoring, support, and friendship.

Thanks to the other members of my committee, Dr. Jessica DeCuir-Gunby, Dr. Andrew McEachin, and Dr. Meg Blanchard, for their constructive feedback and support.

Finally, thanks to my colleagues Daniell DiFrancesca, Dan Spencer, and Ondra Pesout for making this ride all the more enjoyable.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>vi</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>CHAPTER ONE: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Study Purpose</td>
<td>3</td>
</tr>
<tr>
<td>CHAPTER TWO: LITERATURE REVIEW</td>
<td>4</td>
</tr>
<tr>
<td>Self-Regulated Learning</td>
<td>4</td>
</tr>
<tr>
<td>Assumptions of a Self-Regulated Learning Perspective</td>
<td>6</td>
</tr>
<tr>
<td>Self-Regulated Learning</td>
<td>10</td>
</tr>
<tr>
<td>Importance of Self-Regulated Learning</td>
<td>19</td>
</tr>
<tr>
<td>Self-Regulation in Writing</td>
<td>24</td>
</tr>
<tr>
<td>Self-Regulated Learning Interventions</td>
<td>28</td>
</tr>
<tr>
<td>The Self-Regulated Strategy Development (SRSD) Model</td>
<td>32</td>
</tr>
<tr>
<td>Theoretical Perspective</td>
<td>33</td>
</tr>
<tr>
<td>Scope of Model</td>
<td>34</td>
</tr>
<tr>
<td>SRSD Writing Intervention Effects</td>
<td>36</td>
</tr>
<tr>
<td>Feedback</td>
<td>48</td>
</tr>
<tr>
<td>Levels of Feedback</td>
<td>51</td>
</tr>
<tr>
<td>Effects of Feedback</td>
<td>53</td>
</tr>
<tr>
<td>Feedback in Writing</td>
<td>56</td>
</tr>
<tr>
<td>Automated Feedback</td>
<td>61</td>
</tr>
<tr>
<td>Overview of the Present Study</td>
<td>71</td>
</tr>
<tr>
<td>CHAPTER THREE: METHODS</td>
<td>76</td>
</tr>
<tr>
<td>Participants</td>
<td>76</td>
</tr>
<tr>
<td>Study Context: NC Write</td>
<td>79</td>
</tr>
<tr>
<td>Time Frame and Conditions</td>
<td>86</td>
</tr>
<tr>
<td>Instruments and Measures</td>
<td>87</td>
</tr>
<tr>
<td>Procedures</td>
<td>94</td>
</tr>
<tr>
<td>Subjectivity Statement</td>
<td>111</td>
</tr>
<tr>
<td>CHAPTER FOUR: RESULTS</td>
<td>113</td>
</tr>
<tr>
<td>Preliminary Analyses</td>
<td>114</td>
</tr>
<tr>
<td>Main Analyses Overview</td>
<td>116</td>
</tr>
<tr>
<td>RQ1: What changes in students’ writing performance, knowledge, and self-efficacy are associated with NC Write + traditional writing instruction and NC Write + SRSD instruction?</td>
<td>118</td>
</tr>
<tr>
<td>RQ2: What is the shape of growth in students’ writing quality, essay length, and essay elements associated with NC Write + traditional writing instruction and NC Write + SRSD instruction?</td>
<td>131</td>
</tr>
<tr>
<td>RQ3: What are students’ and teachers’ perceptions of the events, value, and strengths and weaknesses associated with NC Write?</td>
<td>143</td>
</tr>
<tr>
<td>CHAPTER FIVE: DISCUSSION</td>
<td>165</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1 Phases and areas for self-regulated learning (Pintrich, 2004)..........................8
Table 2.2 SRSD instructional stages................................................................................35
Table 2.3 Average effect sizes for SRSD studies at different grades, genres, and instructors (Graham & Harris, 2003)...............................................................45
Table 3.1 Participant demographics................................................................................78
Table 4.1 Factor loadings for writing self-efficacy items............................................115
Table 4.2 Results from AGQ confirmatory factor analysis...........................................116
Table 4.3 Means and standard deviations for study variables by condition and time of testing.............................................................................................................117
Table 4.4 Unstandardized coefficients (and standard errors) of multilevel models of writing quality.................................................................................................121
Table 4.5 Unstandardized coefficients (and standard errors) of multilevel models of essay length........................................................................................................123
Table 4.6 Unstandardized coefficients (and standard errors) of multilevel models of essay elements.................................................................................................125
Table 4.7 Unstandardized coefficients (and standard errors) of multilevel models of substantive processes knowledge.................................................................127
Table 4.8 Unstandardized coefficients (and standard errors) of multilevel models of essay element knowledge ..................................................................................129
Table 4.9 Unstandardized coefficients (and standard errors) of multilevel models of writing self-efficacy......................................................................................131
Table 4.10 Unstandardized coefficients (and standard errors) of multilevel growth models of writing quality .........................................................................................136
Table 4.11 Unstandardized coefficients (and standard errors) of multilevel growth models of essay length ..........................................................................................139
Table 4.12 Unstandardized coefficients (and standard errors) of multilevel growth models of essay elements .........................................................................................142
Table 4.13 Student perceptions of NC Write at posttest...............................................144
Table 4.14 Teacher perceptions of NC Write post-intervention....................................145
Table 4.15 Teacher perceptions of how much NC Write assists or impedes students’ writing development..........................................................................................145
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>2 × 2 achievement goal framework (Elliot &amp; MacGregor, 2001)</td>
<td>11</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Student essay annotated with spelling and grammar feedback</td>
<td>81</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Essay total and trait-specific scores</td>
<td>81</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Writing analysis with evaluation and feedback for each trait</td>
<td>82</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Lessons in NC Write</td>
<td>83</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Summary lesson interaction</td>
<td>84</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Drop-down lesson interaction</td>
<td>85</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Framework for deliberate writing practice</td>
<td>151</td>
</tr>
</tbody>
</table>
CHAPTER ONE

Introduction

Researchers have noted the importance of writing well for general academic success, civic involvement, and participation in a global economy (Graham & Perin, 2007b). The Common Core State Standards (CCSS; National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) includes writing standards aimed at ensuring students master a range of skills and applications each year. The CCSS give significant emphasis to not only learning to write but also writing to learn (Graham, McKeown, Kiuahara, & Harris, 2012).

Concerns have been raised about the current quality of students’ writing. The National Assessment of Educational Progress (NAEP) writing assessment was computer administered for the first time in 2011 to students in middle and high school. Results showed the majority of students—approximately 74% in eighth grade and 73% in twelfth grade—performing at the Below Basic and Basic levels in writing, indicating only partial mastery of the knowledge and skills necessary for proficient writing at the respective grades (National Center for Education Statistics, 2012). These concerns and results highlight a clear need for interventions aimed at improving students’ writing proficiency.

Researchers have cited concerns related to the limited amount of writing intervention research, and noted issues with the quality of the extant research in this area (Graham et al., 2012). Bjork, Dunlosky, and Kornell (2013) describe that students are generally not taught how to learn, tend use learning strategies that are ineffective and inefficient, and are not
aware of the limitations associated with such strategies. Though writing instruction in the United States has reflected a shift from a product to a process approach to writing, research suggests that this approach does not provide the amount of support and scaffolding that many students need to develop into effective writers (Graham, Harris, & Mason, 2005; Harris, Graham, & Mason, 2006). Specifically, current process-based approaches to writing instruction, such as the Writers’ Workshop model, do not include the necessary supports to foster the development of self-regulatory skills and motivational beliefs required for proficient writing (Harris, Santangelo, & Graham, 2008).

Graham (2006) has noted that though the positive impact of strategy instruction on writing performance has been convincingly documented classroom application has been limited. Interventions aimed at improving students’ proficiency in any domain need to be sensitive to contextual factors threatening validity or adoption. For instance, sustained deliberate practice is needed to develop expertise in writing (Kellogg & Whiteford, 2009), but this requires teachers to provide continuous writing opportunities and frequent feedback to students—a requirement that is not feasible in most schools given the range of teachers’ instructional responsibilities. Artificial intelligence-based technologies offer the potential to support teachers and provide students with automated scores and feedback to writing compositions. However, the use of these technologies for both assessment and instructional purposes are being debated, and little research has examined classroom applications. There is a need for additional research examining technologies to improve students’ knowledge and abilities as writers that are pedagogically sound and can find support in classrooms across the United States.
Study Purpose

The purpose of the present study is to investigate the effects of a writing intervention that incorporates the self-regulated strategy development (SRSD; Graham & Harris, 1993) model with the automated writing evaluation program NC Write on students’ writing quality, knowledge, and self-efficacy. NC Write provides students and teachers with access to writing prompts, electronic graphic organizers, automated scoring, instant feedback, and interactive lessons. Students in one treatment condition use the writing program and receive traditional writing instruction from their teacher; students in a second treatment condition use the writing program and receive writing instruction based on the SRSD model from their teacher. Students in a comparison condition receive traditional writing instruction from their teacher only. An embedded quasi-experimental mixed methods design is used to examine changes in students’ writing quality, knowledge, and self-efficacy over time as well as students’ and teachers’ perceptions of the writing program.
CHAPTER TWO

Literature Review

Self-Regulated Learning

The capability to regulate our own behavior has been described as possibly the most important quality of humans (Zimmerman, 2000). Self-regulated learning (SRL) considers regulation, and self-regulation, in the full range of learning contexts, such as the classroom, study setting, or workplace (Pintrich, 2000a; Winne & Hadwin, 1998). SRL is generally considered to encompass the cognitive, motivational, behavioral, and contextual components affecting goal-directed learning, and researchers have proposed that SRL mediates the relations among the learner, the environment, and achievement (Pintrich, 2000a). In this regard SRL is inherently contextual (Winne, 2010), and as an active, constructive process SRL involves monitoring, regulation, and control of these components in order to achieve learning goals. SRL is most commonly conceptualized as a cyclical or recursive process, as present performance offers the learner feedback in the form of constantly varying personal, behavioral, and environmental factors that can be considered in order to adjust future performance (Zimmerman, 2000).

As described by researchers SRL is thus a considerable construct; indeed, Paris and Paris (2001) refer to the “broad and indefinite scope of SRL” (p. 89) considered by educational researchers. A number of models of SRL have been proposed in efforts to describe and explain this construct. These models represent a diversity of theoretical perspectives such as cognitive information processing (Winne & Hadwin, 1998), social
cognitive (Pintrich, 2000a; Zimmerman, 2000), and situated (Boekaerts, 2006). Some researchers (e.g., Hadwin, Järvelä, & Miller, 2011; Zeidner, Boekaerts, & Pintrich, 2000) have criticized the profligacy of theoretical perspectives, arguing that this has resulted in confusion about the primary attributes and components of SRL, and has even precluded a common definition of SRL. Regardless, as a theoretical framework SRL has become influential in both psychological and educational research (Azevedo, Johnson, Chauncey, & Burkett, 2010), and has been extended to other fields of study, such as work-related knowledge and skills (e.g., Sitzman & Ely, 2011).

SRL researchers theorize that learners are more likely to increase their knowledge when they approach learning tasks strategically and actively manage their learning (Pintrich, 2000a; Winne & Hadwin, 1998, 2008; Zimmermann, 2000, 2011). Boekaerts (2011) proposed that SRL can be used by learners for task-focused activities such as expanding knowledge and skills, for self-focused activities such as preventing threat and harm, and to protect commitment to the task. SRL is associated with such positive outcomes as academic success, learning motivation, problem solving abilities, and transfer of skills (Pintrich, 2000a; Zimmerman, 2001). Winne (2011) argued that learning is an effortful activity that has the potential for continuous self-regulation because cognitive resources are required to (1) find information that is unknown or cannot be recalled; (2) manage and represent information during learning; and (3) monitor cognition, knowledge, and information through metacognitive processes. Winne (1995) has posited that SRL is a natural process for learners and that SRL is “inherent in learning activities” (Winne, 2011, p. 16).
Assumptions of a Self-Regulated Learning Perspective

Though various models of SRL have been proposed representing a range of theoretical perspectives, they are grounded in shared assumptions about regulation of learning (Pintrich, 2000a). One assumption of SRL models is that learners actively construct meaning, goals, and strategies based on internal (self) and external (environmental) information. This assumption is aligned with constructivist views of learning.

A second assumption of SRL models is that learners have the potential to monitor, regulate, and control aspects of their cognition, motivation, and behavior as well as some aspects of their environments. A number of researchers (e.g., Greene & Azevedo, 2007; Pintrich, 2000a; Zimmerman, 2000) have argued that monitoring is a definitive quality of SRL. This is not to suggest that learners will monitor, regulate, and control these components; in addition to contextual factors there are biological, developmental, and other differences across learners that can constrain the extent to which they are able to self-regulate (Pintrich 2000a, 2004).

Third, models of SRL share the assumption that learners set their own goals, monitor progress towards those goals, and adapt and regulate their cognition, motivation, and behavior to achieve goals. Goals have received considerable examination as they relate to SRL, given that goals serve as standards on which learners’ monitoring of progress and subsequent regulation of cognition, motivation, and behavior is based. Learners hold multiple goals, which can have a number of orientations, may conflict with one another, and may not be aimed towards learning. Researchers have described task-specific goals representing distinct desired outcomes (e.g., to produce an essay that earns the top rubric score) as well as
goal orientations which describe more generally why learners pursue tasks (Dweck & Leggett, 1988; Zimmerman & Kitsantas, 1997). Pintrich (2000a) noted that by examining the “why,” models of SRL incorporate the achievement motivation literature and support a consideration of how the purpose for task engagement affects a learner’s self-regulation. Having goals as standards to support self-evaluation is fundamental to successful SRL, as discrepancies between goals and current performance provide the impetus for strategy change (Hadwin et al., 2011; Muis, 2007).

    The fourth assumption shared by models of SRL is that learners’ self-regulatory behaviors mediate the relations among themselves, the context, and achievement (Pintrich, 2000b, 2004). The implication of this assumption is that learners’ regulation of cognition, motivation, behavior, and the context interact with learners’ individual characteristics, characteristics of the learning context, and learners’ ultimate performance. Models of SRL generally portray direct links between learners’ self-regulatory knowledge and abilities and their achievement (Wolters, Pintrich, & Karabenick, 2003).
A final commonality across models of SRL is that they depict phases or stages of self-regulation that reflect the processes of planning, monitoring, controlling, and reflection (Pintrich, 2004). Models of SRL tend to have a general chronological structure, while also allowing for recursive and cyclical movement through phases. For instance, Pintrich (2000a, 2004) proposed a four phase model consisting of (1) forethought, planning, and activation, (2) monitoring, (3) control, and (4) reaction and reflection, but noted that these phases should not be interpreted to be firmly hierarchical or linear in structure. This model illustrates how

<table>
<thead>
<tr>
<th>Phases</th>
<th>Areas for regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cognition</td>
</tr>
<tr>
<td>Phase 1: Forethought, planning, and activation</td>
<td>Target goal setting</td>
</tr>
<tr>
<td></td>
<td>Prior content knowledge activation</td>
</tr>
<tr>
<td></td>
<td>Metacognitive knowledge activation</td>
</tr>
<tr>
<td>Phase 2: Monitoring</td>
<td>Metacognitive awareness and monitoring of cognition</td>
</tr>
<tr>
<td>Phase 3: Control</td>
<td>Selection and adaptation of cognitive strategies for learning, thinking</td>
</tr>
<tr>
<td>Phase 4: Reaction and reflection</td>
<td>Cognitive judgments Attributions</td>
</tr>
</tbody>
</table>

these phases work across the regulation of cognition, motivation, behavior, and context (see Table 2.1). Wolters et al. (2003) note that while the Pintrich model functions as a theoretical framework for SRL, not all learning provides the opportunity for such explicit SRL as suggested by the model.

Other researches have conceptualized the structure of SRL similarly. For example, Boekaerts (2006) presented a Dual-Processing Self-Regulation model, characterizing SRL as “a set of dynamic, interacting regulation processes” (Boekaerts, 2011, p. 409). Zimmerman’s current model (Zimmerman & Moylan, 2009) comprises forethought, performance, and self-reflection as interrelated, cyclical phases of SRL. Winne and Hadwin’s (1998) model of SRL is described as a “recursive, weakly sequenced system” (p. 281) which includes the stages of (1) task definition, (2) goal setting and planning, (3) enactment, and (4) adaptation; these stages are shown to be influenced by the common facets of conditions, operations, products, evaluation, and standards across stages.

Despite the prevalent focus on the processes of SRL across models, studies have suggested these processes are not so conveniently separated during regulation of learning. Self-report and think-aloud data has shown the interconnections among monitoring, control, and reflection in particular (Pintrich, Wolters, & Baxter, 2000; Pintrich, 2004). As a result of these patterns of findings researchers have recently examined SRL as patterns of events using trace and other data, in an effort to better understand the composition of SRL (c.f., Winne, 2010, 2014).
Self-Regulating Learning

**Regulation of cognition.** The cognitive components of SRL consist of both cognitive strategies and metacognition. Knowledge of cognition and regulation of cognition comprise metacognition (Brown, 1978). Knowledge of cognition includes declarative, procedural, and conditional knowledge (Paris, Cross, & Lipson, 1984), while regulation of cognition includes the subcomponents of planning, monitoring, and evaluation (Schraw & Mossman, 1995). Though experts can develop automaticity in regulating their learning such that SRL is not always deliberate and metacognitive (Winne, 1997), metacognitive monitoring is a primary component of SRL (Winne, 2001), and researchers have demonstrated that metacognition is important for SRL (Hadwin & Webster, 2013; Zimmerman, 2002).

Pintrich (2000) proposed that cognitive planning and activation takes the basic form of goal setting and activation of prior content and metacognitive knowledge. In setting goals for a task, learners may have a mastery or performance orientation to the task (Ames, 1992; Dweck & Leggett, 1988), and may have positive or negative motivation for task engagement, described as an approach or avoidance focus (Elliot, 1997; Elliot & Harackiewicz, 1996). Figure 2.1 presents a 2 × 2 achievement goal framework (Elliot & McGregor, 2001). Recently, Vansteenkiste, Lens, Elliot, Soenens, and Mouratidis (2014) elaborated on this framework—in part to better support examination of the regulation of achievement goals—by incorporating the autonomous and controlled reasons underlying these goals.
**Definition**

<table>
<thead>
<tr>
<th>Valence</th>
<th>Absolute/ intrapersonal (mastery)</th>
<th>Normative (performance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (approaching success)</td>
<td>Mastery-approach goal</td>
<td>Performance-approach goal</td>
</tr>
<tr>
<td>Negative (avoiding failure)</td>
<td>Mastery-avoidance goal</td>
<td>Performance-avoidance goal</td>
</tr>
</tbody>
</table>

*Figure 2.1. 2 × 2 achievement goal framework (Elliot & McGregor, 2001).*

In a self-regulated form, the activation of prior knowledge is a deliberate act of the learner in considering known content and metacognitive strategies that are relevant to a particular task. For instance, a learner could consider that a rehearsal strategy would be sufficient to remember an arbitrary piece of information for a short time, or that an elaboration strategy would be appropriate to encode more complex information.

Cognitive monitoring depends in large part on metacognition. Metacognitive monitoring can be described as an individual’s on-line awareness of comprehension and task performance, seen in the ability to self-test while learning (Schraw & Mossman, 1995). Students may make a variety of monitoring judgments. For example, students may make judgments of their knowledge and retention to determine whether to continue studying for an upcoming test; they also may make judgments of their performance by considering their responses to completed test items. Winne (2011) provided examples of qualities of recalled
knowledge that are likely to be metacognitively monitored, including whether an answer is complete, the degree to which an answer is judged familiar, and an answer’s coherence.

Cognitive control depends on accurate judgments of performance (Boekaerts & Rozendaal, 2010) as monitoring provides the information necessary to support regulation. Thus, while models of SRL generally depict metacognitive monitoring and control as discrete processes, they are conceptualized as being strongly influenced by metacognitive monitoring (e.g., Pintrich, 2000a; Winne & Hadwin, 1998). Conditional knowledge of strategy use is a large part of cognitive control. Students who have procedural and conditional knowledge of a variety of strategies can regulate their cognition by determining which strategy to use—and why—and determining whether to continue use of an effective strategy or discontinue use of an ineffective one. For example, based on monitoring of learning efforts, a student might control and regulate cognition by continuing use of a particular strategy or switch strategies in an effort to improve progress towards the goal.

Finally, cognitive reaction and reflection is grounded in judgments of task performance and the attributions that learners associate with their performance (Pintrich, 2000a). Examinations of Dweck’s (2006) social-cognitive theory of motivation have shown that students with incremental beliefs about intelligence (i.e., beliefs that intelligence is malleable) tend to make more adaptive attributions than those with entity beliefs about intelligence (i.e., beliefs that intelligence is fixed). Attributions can be associated with strategy use, as students with incremental beliefs about intelligence are more likely to attribute failure to a lack of effort or strategy use and to persist by increasing effort or switching strategies, while students with entity beliefs about intelligence are more likely to
attribute failure to a personal characteristic and give up, possibly suffering negative emotions and lowered self-efficacy (e.g., Zimmerman & Kitsantas, 1997). Bandura (1997) defined self-efficacy in terms of beliefs about one’s ability to achieve a goal. Self-efficacy is context-specific and thus the level, generality, and strength of self-efficacy can vary by task (Zimmerman, 1995).

**Regulation of motivational beliefs, affect, and emotions.** Learners can additionally regulate their motivational beliefs, affect, and emotions. Researchers have identified self-efficacy beliefs, goal orientations, task interest/value beliefs, and outcome attributions as components of achievement motivation. However, SRL researchers have addressed the role motivation plays in SRL to varying degrees. For example, Pintrich’s (2000a) model is grounded in motivational as well as cognitive processes, while motivational regulation is ambiguous in Winne and Hadwin’s (1998) model of SRL. Some authors have thus argued that the regulation of motivation has been underemphasized in the SRL literature (Pintrich, 2000a; Wolters, 2003), and the scarcity of emotion regulation research has also been criticized (Boekearts, 2011). Boekearts (2011) defines emotion regulation as the “capacity to understand one’s own emotions and their expression, as well as to the capacity to bring order by modifying or tempering aspects of the emotional experience, when it interferes with the pursuit of important goals and with social interaction” (p. 415). Students can regulate emotions in order to achieve learning goals (Pekrun, Frenzel, Goetz, & Perry, 2007) and empirical research has documented the link between students’ emotions and academic performance (Pekrun, Elliot, & Meyer, 2009).
Motivational planning and activation involves learners’ judgments and beliefs about the learning task and themselves. Students can consider task-relevant self-efficacy beliefs, their expectations of and value associated with the task (c.f., Eccles et al., 1983; Wigfield & Eccles, 2000; Wigfield, Tonks, & Eccles, 2004), perceptions of task difficulty, and any relevant affective and emotional factors, such as interest or anxiety, as these relate to their understanding of the task.

Monitoring of motivational beliefs, affect, and emotions involves learners’ awareness of these factors in the context of task goals. During task enactment learners can be self-aware of current levels of efficacy, consider present perceptions of task value and difficulty, and attend to prevailing levels of affective and emotional components, such as interest, disappointment, pride, etc., as these relate to the learning task. These evaluations may suggest realistic present states given the task context, or may reveal disconnects between these factors and learning goals.

Monitoring of motivational beliefs, affect, and emotions may prompt learners to exert control over these factors in the form of affording to maintain present states or striving to better adapt their motivations and emotions to the task in order to increase the likelihood of achieving learning goals. Wolters (2003) presented a review of the motivational regulation literature. Learners commonly regulate their motivation by self-administering consequences in the form of self-imposed rewards or punishments contingent on performance. Students have been known to use goal-oriented self-talk strategies (Wolters, 1998, 1999), and self-talk can also be used to temper anxiety. In order to control self-efficacy, learners can use efficacy self-talk, set proximal goals, or self-induce anxiety to stimulate motivation by convincing
themselves they are not prepared for the demands of a task using defensive pessimism. Students may also use maladaptive strategies to control motivation, such as self-handicapping, by decreasing effort or procrastinating (Wolters, 2003). Emotional control can be seen in the form of adaptive (focus on the positive, reduce tension) and maladaptive (worry, ignore the problem) coping strategies. Learners can control emotions and negative affect by suppressing or expressing emotions, self-denying information or self-distracting away from the task causing the emotional reaction, re-appraising the task, or acquiring and providing social supports (Boekaerts, 2011).

Motivational reaction and reflection involves learners’ emotional reactions to the results of task performance as well as attributions for success or failure. Attributions associated with task performance can be associated with a range of complex emotions. Pekrun et al. (2007) define achievement emotions as emotions directly associated with an achievement activity or outcome of the activity. The authors classify achievement emotions according to whether they are positive (pleasant) or negative (unpleasant) and further to the degree of activation. For example, in reacting to a passing grade on an exam, a student might experience an activating positive emotion (pride), or a deactivating positive emotion (relief). By controlling their attributions, learners can maintain efficacy and motivation for subsequent tasks (Pintrich, 2000a).

**Regulation of behavior.** Learners can also regulate their own behavior. Many models of SRL (e.g., Hadwin, Järvelä, & Miller, 2011; Pintrich, 2000a; Zimmerman, 2000) conceptualize behavior as an aspect of the individual and thus of self-regulation; however, there are some models (e.g., Winne & Hadwin, 1998) that do not conceptualize behavioral
regulation as directly related to regulation of the self. Therefore, regulation of behavior is not consistently included across models of SRL. Pintrich (2000a) argued that since behavior can be observed, monitored, and controlled, it is self-regulatory in nature.

Behavioral planning includes planning for the time and effort one will devote to a task (Pintrich, 2000a). Learners can use time management strategies to allocate time across tasks, and during such planning learners can also consider how they intend to allocate their effort across tasks. Learners may plan for self-observation of their own behavior in order to have a strategy in place for collecting behavioral data during task enactment and to support behavioral monitoring and control.

Behavioral monitoring can be in the form of record-keeping (e.g., logs, diaries), tracking of tasks completed, or output. Self-recordings (Zimmerman, 2000) can be examined in the context of learning goals to evaluate patterns in time and effort expense as well as progress toward goals. Zimmerman (2000) cautioned that complex performances can be quite difficult to monitor accurately due to the amount of information available. Thus it is important to monitor the right data—and interpret these data accurately—to support regulation of strategy use or effort.

A range of behavioral control strategies are discussed in the literature. Based on monitoring results learners can reallocate time across tasks and regulate effort to better fit task contexts. Students may seek help from teachers, peers, or others when they become aware they are having difficulty with a learning task. Learners can use attention focusing strategies to maintain focus on a task and reduce distractions. Pintrich (2000a) emphasized the link between the regulation of motivational beliefs, affect, and emotions, and the
regulation of behavior, and this relation has also been documented empirically (e.g., Marchand & Skinner, 2007; Thompson & Richardson, 2001; Wigfield, Eccles, Schiefele, Roeser, & Davis-Kean, 2006). Both adaptive and maladaptive motivational strategies impact a learner’s effort and thus have implications for behavioral regulation. For example, students can regulate their behavior by employing a self-handicapping behavior such as procrastination which will regulate behavior (through decreased effort) and motivation (by protecting self-worth).

Behavioral reaction and reflection is most accurately characterized as a cognitive rather than behavioral activity, and reflecting on behavior is an important part of SRL (Pintrich, 2000a). Behavioral reaction and reflection is grounded in judgments of task performance, considering behavioral aspects such as time management and effort expense. These judgments have implications for future task enactment, as learners make decisions about which tasks to engage in and how to approach those tasks (Pintrich, 2000a).

**Regulation of context.** Finally, learners can regulate those aspects of the external environment and task that are within their control. Pintrich (2000a) explains, “It is the self or person who is acting on the context and attempting to change it as well as adapt to it that makes attempts to regulate the context a part of self-regulated learning” (p. 456). Previous research has documented the impact of context on learning (c.f., Perry, Turner, & Meyer, 2006). Specifically, researchers have shown that educational contexts influence students’ self-regulatory strategy use and effort (Paris & Paris, 2001). The social context determines whether learners might engage in co-regulated or socially shared regulation of learning (c.f., Hadwin, Järvelä, & Miller, 2011).
In formal educational contexts, contextual forethought, planning, and activation involve students’ perceptions of teacher expectations and practices, as well as norms associated with the classroom context and task. Pintrich (2000a) notes that such perceptions may be inaccurate, grounded in stereotypes or other beliefs. Outside of the classroom environment students have greater opportunity to structure their learning contexts—and likely more autonomy to do so than in formal educational settings—and could, for example, organize a study space free from distractions in their homes.

Contextual monitoring involves awareness of aspects of the context relevant to learning tasks. In a formal educational context, it is salient for students to understand classroom rules and policies, particularly those related to task evaluation or reporting of performance such as assessment and grading practices. In informal or individual learning environments, students can consider environmental characteristics as related to task goals. For example, a student may develop an understanding that the local library is not an ideal study setting due to high traffic or noise levels.

Students can afford to control the task and context based on monitoring efforts. Pintrich (2000a) suggests that contextual factors are more difficult to regulate than personal factors. However, in classroom settings students can be involved to varying degrees in the design and administration of rules, policies, and individual tasks so they have opportunities to regulate these aspects of the context. The degree to which a student is able to control or regulate the context depends on the beliefs and pedagogy of the teacher, with a student-centered classroom generally offering greater opportunities for student control than a traditional classroom (Pintrich 2000a, 2004). Co-regulation may occur when peers encourage
one another to use a regulatory process or strategy (Hadwin, Järvelä, & Miller, 2011), such as reminding each other to self-evaluate their work using a rubric. In individual learning environments, students have greater autonomy to manipulate the environment based on monitoring efforts so that it is conducive to the pursuit of learning goals. This could include restructuring the environment to better organize materials needed for a task or to adjust or relocate the environment to reduce distractions.

Contextual reflection involves assessment of the learning environment as related to task performance, in addition to general evaluations of environmental conditions (Pintrich, 2000a). In formal educational contexts, students can be involved in periodic reflections on class—and classroom—success and areas for improvement, and may offer unsolicited feedback. Learners may also critique the private settings in which they study or complete assignments in light of their performance to determine what might be adjusted during future planning.

**Importance of Self-Regulating Learning**

Many students have been found to have sub-optimal levels of self-regulation (Azevedo & Cromley, 2004). Students frequently need to learn about topics that do not support ideal levels of motivation (Hidi & Harackiewicz, 2000). An increasing amount of learning occurs in informal learning environments, requiring individuals to take more responsibility for their own learning (Bjork, Dunlosky, & Kornel, 2013; Schunk & Zimmerman, 1998). For example, online learning environments tend to comprise fewer forms of reinforcement and redirection to learning goals than more formal environments where teachers and peers are present (Delen, Liew, & Willson, 2014). SRL is critical to learn
complex topics in online environments (Jacobsen, 2008), and “an important survival tool” (Bjork, Dunlosky, & Kornel, 2013, p. 418) for success in modern learning contexts.

The construct of SRL has received extensive research by respected educational psychologists, and empirical results have supported the evolution of theory and been of practical use to educators and learners (Boekaerts, 1999). Studies showing that students’ academic achievement could not be fully explained by their skills and abilities alone piqued interest in examining SRL in educational contexts to explain achievement differences (Zimmerman, 2001). Researchers have come to understand the importance of SRL by examining the impacts of SRL on academic outcomes, strategy use, achievement motivation, as well as by examining the development of SRL.

A wide range of studies have documented the impact of SRL on academic achievement (c.f., Boekaerts, Pintrich, & Zeidner, 2005; Dignath & Büttner, 2008; Dignath, Buettner, & Langfeldt, 2008; Efklides, 2011). SRL has also been associated with a variety of academic and learning outcomes. Self-regulated learners have greater engagement in learning and information retention (Delen et al., 2014). SRL has been associated with deep conceptual understanding due to learners’ cognitive and motivational investment in learning (Zohar & Ben-David, 2009). Some research has suggested there may be a stronger impact of SRL on achievement for students with learning disabilities (Graham & Perin, 2007a; Ruban, McCoach, McGuire, & Reis, 2003). Recent co-regulated learning research has suggested potential benefits for a learner’s SRL activity and capacity as a result of regulating learning with peers and in larger social environments, though limited empirical work has been done in this area (Hadwin et al., 2011).
SRL has also been linked to strategy use. Self-regulated learners set hierarchal goals related to both process and product (Zimmerman, 2000). Self-regulated learners, like experts, spend a greater amount of time planning, which results in their higher achievement (Zimmerman, 2008). Students who plan for and implement a strategy are more likely to attribute failure to strategy use rather than to their own ability (Cleary & Zimmerman, 2001; Zimmerman & Kitasantas, 1997). The use of time management strategies has been observed with high-achieving self-regulated learners (Zimmerman & Martinez-Pons, 1986). The effective use of monitoring strategies can allow for improved SRL (Zimmerman & Moylan, 2009) and improved performance (Thiede, Anderson, & Therriault, 2003; Nietfeld, Cao, & Osborne, 2006). Researchers have also demonstrated the impact of assessment opportunities on self-monitoring and subsequent SRL (Thiede, Redford, Wiley, & Griffin, 2012). Interest enhancement strategies used by self-regulated learners have been associated with greater effort, use of metacognitive strategies, and task value (Wolters, 1999; Wolters & Rosenthal, 2000). High-achieving, mastery-oriented self-regulated learners are most likely to demonstrate adaptive help-seeking behaviors (Karabenick & Newman, 2006; Roll, Aleven, McLaren, & Koedinger, 2011; Ryan, Patrick, & Shim, 2005). Reflecting on general findings related to SRL and strategy use, Bjork et al. (2013) criticize the over-appreciation of individual differences and under-appreciation of the role of effort and practice.

Researchers have also examined the relations between SRL and motivation. There is a strong link between self-efficacy and SRL (Ferla, Valcke, & Schuyten, 2008; Pintrich, 2004). Self-efficacy has been shown to predict achievement (Ferla et al., 2008; Robbins et al., 2004), motivation, and academic development (Zimmerman & Cleary, 2006). Self-efficacy
beliefs have been associated with strategic planning (Cleary & Zimmerman, 2001; Zimmerman & Martinez-Pons, 1990). Students with higher levels of self-efficacy tend to set higher and more specific goals (Cleary & Zimmerman, 2001), and self-efficacy has been shown to have a causal effect on goal setting (Locke & Latham, 1990; Zimmerman & Bandura, 1994). Furthermore, higher levels of self-efficacy are related to task choice, energy expenditure, and persistence, and negatively related to procrastination (Wolters, 2003; Zimmerman & Cleary, 2006). The relations between self-efficacy and evaluations of performance or progress have been demonstrated empirically (e.g., Zimmerman & Kitsantas, 1999). Self-efficacy has been conceptualized as an antecedent to achievement goals (Elliot & Church, 1997), and reciprocally related to causal attributions (Zimmerman & Cleary, 2006).

The adoption of mastery goals has been associated with a wide range of self-regulatory and learning benefits: cognitive organization and elaboration strategy use, comprehension and progress monitoring, intrinsic motivation, positive self-efficacy, adaptive attributions, persistence and recovery from setbacks, achievement, and overall greater self-regulation (Bernacki, Byrnes, & Cromely, 2012; Grant & Dweck, 2003; Elliot, 1999; Senko, Hulleman, & Harackiewicz, 2011; Wigfield et al., 2006; Wolters, 2004). A performance-approach orientation is generally understood as more adaptive than a performance-avoid orientation (Friedel, Cortina, Turner, & Midgley, 2007), though findings examining performance goals have been mixed. Performance goals in general have been associated with the use of self-handicapping strategies (Urdan, 2004); however, the adoption of performance-approach goals has been shown to be related to some desirable traits, including self-regulation and positive self-efficacy (Wolters, Yu, & Pintrich, 1996). As additional
motivational components, task value and interest can prompt student engagement and academic outcomes, and have been shown to be related to the use of SRL (Pintrich & Zusho, 2002; Schunk, Pintrich, & Meece, 2008).

Researchers have also examined emotions as related to SRL. Positive affect is related to effort and strategy use (Linnenbrink & Pintrich, 2002, 2004), while negative activity and outcome emotions have been shown to negatively predict SRL (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). Positive emotions are an important determinant of students’ SRL and academic success (Mega, Ronconi, & De Beni, 2014; Pekrun et al, 2009). Emotional control can be used to reduce test anxiety, though studies examining the relations between motivational control strategies and academic achievement have shown mixed results (Zeidner, 1998).

The development of SRL has received limited examination. Early evidence suggested that young students have difficulty regulating their learning, and showed developmental shifts in SRL knowledge and abilities throughout the elementary years (Paris & Newman, 1990). However, additional evidence has suggested that even young students are capable of regulating their learning to some extent (Perry, Phillips, & Dowler, 2004; Perry, VandeKamp, Mercer, & Nordby, 2002). Pintrich and Zusho (2002) discussed how development of SRL is influenced by cognitive and motivational factors germane to the learning context. Development of students’ content and metacognitive knowledge, proficiency with strategy use, epistemological beliefs, and knowledge of self encourages SRL; however, shifting beliefs about intelligence, goal orientations, task values, and interest may at the same time discourage SRL. Pintrich and Zusho (2002) explain, “these two general developmental trends
suggest the paradoxical conclusion that just as students are becoming cognitively able to self-regulate, their motivational beliefs and some aspects of the school context conspire to deflect them away from self-regulation activities” (p. 278).

Self-Regulation in Writing

The act of writing has historically been conceptualized as a recursive process that requires self-regulation, in particular cognitive monitoring (c.f., Bereiter & Scardamalia, 1987; Flower & Hayes, 1980). However, early models of writing have focused on describing cognitive processes as they relate to writing competence (Zimmerman & Kitsantas, 2007). Considering the importance of SRL in translating competence into performance, Zimmerman and Kitsantas (2007) characterize writing as “a social cognitive process wherein writers must be willing to devote personal time and effort to revise text drafts until they communicate effectively” (p. 53). The high cognitive demands and complexity of writing require that students have self-regulatory skills to write well (Graham & Harris, 2000; Zimmerman & Kitsantas, 2007).

Self-regulation of writing compromises self-initiated thoughts, feelings, and actions used to achieve writing goals, including improving the quality of the writing process and writing skills as well as the products of writing (Schunk & Zimmerman, 1994). Zimmerman and Risemberg (1997) presented a detailed model of self-regulated writing. The authors explained that a writer may regulate cognition and motivation using planning, goal setting, and self-evaluation strategies. Personal processes can be regulated through the use of cognitive strategies, such as pre-writing and outlining, and the use of mental imagery, as might be used to prompt creativity. A writer may regulate his or her behavior by monitoring
writing output or quality against a previously defined goal or by self-imposing consequences, perhaps in the form of a reward after completion of part of the task, be it a writing assignment, certain number of pages, or chapter. Additionally, a writer may use self-verbalizations in the form of self-questioning or produce self-instructions to refer to during the writing process. Finally, a writer may regulate the context in the form of his or her environment by manipulating the physical space or making available resources either to incorporate or to employ as models. Writers monitor and control environmental, behavioral, and personal processes based on perceptions of feedback and motivational aspects relevant to the writing task (Zimmerman & Risemberg, 1997). Researchers have provided numerous examples of writing self-regulation strategies used and discussed by famous novelists, poets, and professional writers (see Zimmerman & Kitsantas, 2007; Zimmerman & Risemberg, 1997). Furthermore, studies of professional writers have shown expert writing performance to include SRL behaviors, such as prewriting and revision strategies to manage cognitive load, environmental structuring to reduce distraction, and emotional regulation to support persistence (Kellogg, 2006).

It thus comes as no surprise that researchers have found links between SRL and effective writing (e.g., Graham & Harris, 2000; Zimmerman & Kitsantas, 1999). Self-monitoring is critical to write effectively (Graham et al., 2005; Winne & Hadwin, 1998), and improvements in writing self-monitoring have been associated with improvements in writing quality (Cho, Cho, & Hacker, 2010; Winne & Hadwin, 1998). Students who set specific goals for their writing produce more effective writing than those who set general goals or do not set goals (Ferretti, MacArthur, & Dowdy, 2000; Graham, McArthur, & Schwartz, 1995).
The use of mental imagery has shown positive impacts on writing quality (Jampole, Matthews, & Konopak, 1994). Hacker, Keener, and Kircher (2009) emphasized the importance of metacognitive monitoring and control for writing, arguing that writing is applied metacognition. Empirical research has demonstrated the impact of metacognition on writing performance (Schunk & Zimmerman, 2007; Zimmerman & Bandura, 1994). Observational learning about writing prompts SRL strategy use (Braaksma, Rijkaarsdam, Van den Bergh, & van Hout-Wolters, 2004), and has been shown to have a positive impact on writing performance (Zimmerman & Kitsantas, 2002).

Self-efficacy has also been shown to be a powerful predictor of writing performance. Writing self-efficacy mediates the relations between writing apprehension and achievement, and explains variance in writing performance controlling for initial writing performance, grade, and gender (Pajares, Britner, & Valiante, 2000; Zimmerman & Bandura, 1994). On average, females are found to have higher writing self-efficacy than males (Andrade, Wang, Du, & Akawi, 2009; Pajares, Miller, & Johnson, 1999). Self-efficacy is thought to be closely linked to interest (Ainley, Hidi, & Berndorff, 2002; Zimmerman & Kitsantas, 1999). Researchers have examined self-efficacy specific to the self-regulation of writing. Findings have shown students’ beliefs about their ability to use writing strategies are related to strategy use, adaptive attributions, and writing performance (Schunk & Zimmerman, 1994); however, self-efficacy for writing self-regulation does not explain additional variance in writing performance above writing self-efficacy (Pajares & Valiante, 2001). Bruning, Dempsey, Kauffman, McKim, & Zumbrunn (2013) provided evidence for a multifactor
model of writing self-efficacy, comprised of self-efficacy for ideation, conventions, and writing self-regulation.

Researchers have documented connections between both goal setting strategies and goal orientation and writing self-efficacy. For example, Schunk and Schwartz (1993) found that students who set process goals aimed at effective strategy use reported higher levels of writing self-efficacy than students who set general goals. Students who additionally received feedback reported the highest levels of self-efficacy at posttest (Schunk & Schwartz, 1993). Pajares, Britner, and Valiante (2000) found that performance-approach goals were associated with general writing self-efficacy and self-efficacy related to using self-regulatory strategies for 7th and 8th grade students but not younger children, leading the authors to conclude that performance-approach goals may be more facilitative as students progress through middle school.

Finally, despite the limited longitudinal research examining the development of SRL skills and abilities in writing, the available evidence suggests that writers’ SRL improves with age and task experience (Graham & Harris, 2000; Pintrich & Zusho, 2002). For example, researchers have documented improvements in writing planning (Bereiter & Scardamalia, 1987) and conceptualizations of writing (Graham, Schwartz, & MacArthur, 1993) over time.

A few recent meta-analyses have examined experimental and quasi-experimental writing interventions and offer some generalizations about effective writing instruction as related to SRL. Graham and Perin (2007a) examined writing interventions conducted with grade 4–12 students and reported positive effects on writing quality associated with the study of models (ES=0.25), product goals (ES=0.70), strategy instruction (ES=0.82), and SRL-
based strategy instruction (ES=1.14), among other treatments. Graham et al. (2012) examined writing interventions conducted with elementary grade students and reported positive effects on writing quality associated with creativity/imagery instruction (ES=0.70), product goals (ES=0.76), and strategy instruction (ES=1.02), among others. The authors determined that the addition of SRL instruction to strategy instruction further increased writing quality (ES=0.50). Graham and Harris (2000) proposed that overall, teaching students SRL strategies improves their writing performance.

**Self-Regulated Learning Interventions**

Given the diverse representation of theoretical perspectives in models of SRL, Paris and Paris (2001) argued that theoretical foundations of SRL are less important to educators than practical applications for teaching and learning. This can be seen in the breadth of interventions aimed at improving students’ SRL. The majority of the SRL research has been conducted with students in the upper elementary grades through college (Perry et al., 2004), thought recent intervention studies have shown that elementary grade students can be taught to effectively regulate their learning (Dignath et al., 2008).

Researchers have used strategy training interventions to improve students’ SRL and performance (Azevedo & Cromley, 2004; Nietfeld & Schraw, 2002). The incorporation of SRL strategies enhances the effects of cognitive strategy instruction on learning outcomes (Schünemann, Spörer, & Brunstein, 2013; Spörer & Schünemann, 2014). Interventions aimed at improving students’ strategy use, self-monitoring and -evaluation, and goal orientations have shown positive effects on SRL (Schunk, 2005). Recent interventions to improve students’ SRL have shifted away from an emphasis on domain-specific strategy
instruction to more balanced models that situate strategy instruction that includes procedural and conditional knowledge about strategies in learning contexts conducive to the development of SRL (Boekaerts & Corno, 2005; Lin, 2001). For example, Spörer and Schünemann (2014) showed positive effects of strategy implementation and outcome regulation procedures in a reciprocal teaching context. In the context of an elementary writing program, Harris, Graham, and Mason (2006) documented impacts of a strategy training intervention that included self-regulatory knowledge and process instruction as well as procedures to support students’ motivation on students’ writing performance and knowledge.

Recent efforts have focused on designing computer-based learning environments (CBLEs) that scaffold students’ SRL and metacognition (Azevedo, Johnson, Chauncey, & Burkett, 2010; Roscoe, Segedy, Sulcer, Jeong, & Biswas, 2013), based on the importance of incorporating conceptual, procedural, metacognitive, and strategic scaffolds to help students achieve learning goals (Azevedo & Hadwin, 2005). When sufficient scaffolds are not in place in CBLEs students have difficulty regulating their learning, which ultimately constrains their conceptual understanding (Azevedo, Moos, Greene, Winters, & Cromley, 2008). The freedom to control learning that CBLEs offer students is a double-edged sword; while it helps to maintain student engagement in complex tasks, students may not engage with aspects of CBLEs designed to support SRL (for a review see Roscoe et al, 2013). Students may game CBLEs intentionally or unintentionally, completing learning tasks by circumventing the goals intended by the developer.
Despite the design challenges, researchers have successfully supported students’ SRL and performance in CBLEs using a range of interventions such as interactive videos (Delen et al., 2014), open-ended teachable agent systems (Leelawong & Biswas, 2008), and interim testing experiences (Soderstrom & Bjork, 2014). The extant research examining SRL in CBLEs is focused at the undergraduate level, though there are an increasing number of CBLEs being designed to support younger students’ SRL. Tsai, Shen, and Fan (2013) examined online SRL studies published in selected journals from 2003-2012 and found that only 10.8% sampled elementary and middle school students. Additionally, the majority of the research in this field has been purely quantitative. Tsai et al (2013) reported that of the same studies, only 4.3% were qualitative and 19.6% were mixed-methods, suggesting a need for more diverse methodologies to investigate how to improve students’ SRL in the context of CBLEs.

Three recent meta-analyses have described the overall effectiveness of SRL interventions. Dignath et al. (2008) examined 48 studies conducted in elementary schools and found an overall positive effect of SRL interventions on academic performance, strategy use, and motivational outcomes, based on an overall weighted average effect size (ES) of 0.69. The authors concluded that interventions for elementary school students had the greatest impact on these outcomes when they were based on social-cognitive theories, combined different types of strategy instruction (e.g., metacognitive and motivational), and provided students with procedural and conditional knowledge about strategies as well as feedback about strategy use.
Dignath and Büttner (2008) examined 74 studies conducted in both elementary and middle schools and calculated an overall weighted average effect size of SRL interventions on academic performance, strategy use, and motivational outcomes of 0.69. Interventions for elementary school students had the greatest impact on academic performance and strategy when they were based on social-cognitive theories and emphasized motivational aspects of SRL, while interventions for middle school students had the greatest impact on academic performance and strategy use when they were based on metacognitive theories. Interventions had greater effects for strategy use when they supported students’ procedural and conditional knowledge of strategies. Dignath and Büttner (2008) found weighted effect sizes for cognitive, metacognitive, and motivational strategy use were higher in middle school (0.88) than elementary school (0.72) settings. The authors also found that SRL interventions achieved lower effects in reading/writing contexts compared to math and other subjects. These findings may be explained by younger students’ difficulty applying metacognitive strategies to cognitively demanding contexts like reading and writing in which they tend to be inexperienced (Alexander, Graham, & Harris, 1998).

Finally, Donker, De Boer, Kostons, van Ewijk, & Van der Werf (2014) conducted a meta-analysis of 58 studies carried out in elementary and middle schools to determine the effect of SRL strategy instruction on academic performance, calculating an average mean effect size of 0.66. The authors found the highest effects in writing, regardless of the strategy focus of the intervention, and determined that combining strategies in interventions almost always improved performance. Additionally, Donker et al. (2014) reported that the addition of metacognitive knowledge to interventions increased academic performance, and found
strategies focused on planning and task value overall effective, hypothesizing that such interventions addressed motivational aspects of learning.

A short list of best practices can thus be drawn from the SRL intervention literature. Interventions should combine cognitive, metacognitive, and motivational strategy training (Boekaerts & Corno, 2005; Dignath et al., 2008). They should include appropriate scaffolds to support SRL, in particular in CBLEs, and should provide students with feedback about learning strategies. Finally, interventions should offer opportunities for students to achieve multiple goals in the context of complex tasks (Perry et al., 2002, 2004).

**The Self-Regulated Strategy Development (SRSD) Model**

The Self-Regulated Strategy Development (SRSD) model (Graham & Harris, 1993) was designed to help students develop both the cognitive and self-regulatory skills necessary for proficient performance in an academic domain. SRSD acknowledges skill development as a complex task that affects students’ content knowledge, strategic knowledge, and motivation, in line with current theories of learning (Alexander, Graham, & Harris, 1998). Based on this premise, SRSD reflects many of the best practices drawn from the SRL intervention literature, in that it combines different types of strategy instruction to support students’ development of knowledge and SRL procedures in the context of complex academic tasks. SRSD has been used in a number of content areas, including math, reading, and writing (Wong, Harris, Graham, & Butler, 2003). The SRSD writing model is the focus of the remainder of this section.

SRSD is grounded in the idea that self-regulation plays an important role in students’ academic success, and that students should receive explicit instruction related to self-
regulation, and be supported in the development of their self-regulatory knowledge and abilities throughout writing instruction, in order to develop the ability to regulate their learning independently. Additionally, the SRSD model is designed to address students’ motivational beliefs, by supporting their self-efficacy and encouraging adaptive attributions. Instructional application of the SRSD writing model emphasizes active learning supported by appropriate scaffolds, valuing of writing and the relations between effort and writing quality, and an instructional environment that prioritizes writing by providing numerous non-threatening writing opportunities, modeling adaptive beliefs, and providing students with feedback. SRSD is characterized as responsive instruction (Harris et al., 2008b), as strategies and instructional components are intended to be adapted to meet students’ learning needs. Additionally, SRSD is designed to be criterion-based, in that students demonstrate proficiency with strategy use and self-regulatory procedures at each stage of instruction before proceeding to the next.

**Theoretical Perspective**

The original conception of the SRSD model was largely informed by the constructivism literature, in particular research related to strategy instruction and acquisition. Specifically, development of the SRSD model was influenced by Meichenbaum’s (1977) cognitive-behavior modification model, Vygotsky (1978) and colleagues’ social-cognitive research, Deshler (1981) and colleagues’ investigations of the strategy acquisition process of students with learning disabilities, and Brown (1981) and colleagues’ research examining strategy training to support SRL. As the model was refined over time, it came to reflect a range of theoretical approaches, in part to support a blending of evidence-based practices
(Harris et al., 2008b). As Harris et al. (2008b) explain, “an underlying premise of SRSD is the necessity to integrate diverse, validated pedagogical approaches, regardless of whether their theoretical bases are viewed by some as discordant” (p. 399). SRSD has thus evolved to reflect a wide variety of theoretical perspectives, including social cognitive, information processing, and sociocultural, and continues to be informed by emerging research related to instructional practices, the development of academic knowledge and skills, learner characteristics, writing expertise, and self-regulation (Harris et al., 2008b).

**Scope of Model**

The SRSD model for writing includes (1) explicit instruction in narrative, expository, and opinion/persuasive writing purposes presented over six stages of instruction (described below); (2) explicit instruction in SRL strategies; and (3) development of positive attitudes towards writing and writing self-efficacy (Harris, Graham, & Mason, 2003; Santangelo, Harris, & Graham, 2007). The self-regulatory skills targeted by SRSD instruction include goal setting, self-monitoring, self-assessment, self-instruction, self-reinforcement, imagery, and managing the writing environment (Harris et al., 2008b). The emphasis on self-regulatory skills provides students with tools to help manage their use of writing strategies and the writing task and allows students to collect evidence of their own writing growth (Harris et al, 2006).

The six stages of SRSD instruction provide a set of general guidelines for instruction and the stages can be modified or revisited as needed (Harris et al., 2003; Harris et al., 2008b). For instance, researchers have found success modifying the SRSD writing model to better meet the needs of students with behavioral and emotional difficulties by allowing...
increased time to provide background knowledge and teach the strategy prior to modeling strategy use (Lane et al., 2008). The SRSD instructional stages are presented in Table 2.2. Scaffolded support is faded across the stages of instruction in a gradual release of responsibility, so that students apply a strategy independently by the end of instruction.

Table 2.2

<table>
<thead>
<tr>
<th>Stage</th>
<th>SRSD Instructional Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>Develop and activate background knowledge</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Discuss it</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Model it</td>
</tr>
<tr>
<td>Stage 4</td>
<td>Memorize it</td>
</tr>
<tr>
<td>Stage 5</td>
<td>Support it</td>
</tr>
<tr>
<td>Stage 6</td>
<td>Independent performance</td>
</tr>
</tbody>
</table>

For each strategy (e.g., planning, drafting, revising, sentence generation, sentence combining), students are first taught the background knowledge needed for effective strategy use. Instruction at this stage also addresses the use of positive self-statements (“I can be successful at this task by using the strategy”) to support effective strategy use. Students are next taught the strategy itself along with a mnemonic to support learning the strategy, and provided with information related to why the strategy is effective. Additionally, instruction addresses the connection between effort and successfully strategy use to encourage adaptive attributions. Students can set goals for strategy use based on evaluations of current writing performance. The strategy is then modeled for students, and information about the benefits and limitations of the strategy is provided. The use of self-instructions including those related to planning (“What do I need to do first?”), self-evaluation (“Did I use the strategy
appropriately?"), and self-control ("I can do this if I complete one part at a time") are also modeled, and students can develop personalized self-instructions. The goal of the next stage of instruction is for students to commit the strategy to memory, both in terms of the strategy process and self-statements they have associated with it. Next, students practice use of the strategy with faded support until they are able to demonstrate proficient strategy use. The focus of the final stage of instruction, where students are using the strategy independently, is on evaluation of strategy use and writing performance. SRSD instruction encourages the maintenance and generalization of writing and self-regulatory strategies in a number of ways, such as by having students identify opportunities to use strategies in other settings and evaluate the success of strategy use (Harris et al., 2008b).

**SRSD Writing Intervention Effects**

SRSD writing intervention studies have employed group designs, including experimental and quasi-experimental designs, and single-subject designs, in particular multiple baseline across participants designs. Intervention effects have been most frequently calculated at posttest; however, a minority of SRSD studies also report effects at midtest, maintenance, and/or include generalization effects, based on measures of transfer across contexts or genres. SRSD writing intervention studies commonly examine outcome measures of writing quality, length, elements, and story grammar associated with treatments (Graham & Harris, 2003). Measures of writing quality address overall composition value; students are commonly provided a choice from two equivalent writing prompts to address motivational concerns. While a holistic measure of writing quality (0–7 or 0–8 scale) is most commonly used, raters in SRSD research tend not to be provided with rubrics and are trained solely
using grade-specific low, middle, and high anchor papers. Quality scores are derived as the average of two raters’ scores. Length is measured by the number of words in a composition; composing time may also be captured. Elements comprise genre elements in story writing (e.g., main characters, characters’ actions to achieve goals, consequences of actions, etc.), and basic elements of persuasive writing, including the premise, reason, examples, and conclusion. Elements are scored awarding a 0 if the element is absent and 1 if it is present, with the exception of the persuasive writing elements of reasons and examples, which are scored awarding a 1 for each unique reason and example present. Story grammar has been measured similarly to elements, assessing both the presence and overall quality of elements.

A wide range of additional measures and outcomes have been examined in SRSD research. Investigations of planning have involved the examination of any notes students produce prior to writing and measures of planning time prior to the start of composing. Writing time has also been collected. Graham (2006) reports a wide range of measures of revisions used in SRSD research, including surface-level and non-surface-level revisions; meaning-changing and -preserving revisions; meaning non-surface-level revisions; total revisions; revising time; additions, deletions, and substitutions; and errors detected and corrected. Researchers have examined the number of transition words used in essays, vocabulary diversity, and the ratio and variety of modifiers. Some SRSD studies have included measures of mechanics, such as spelling, punctuation, and conventions; however, there seems to be a trend away from examining such measures, and in more recent SRSD intervention studies writing examiners have provided students with word spellings on request.
Writing knowledge surveys have been used as pre- and post-measures of students’ knowledge about good writing and the behaviors of good and poor writers as well as knowledge about planning. Additional open-ended items have been used to assess students’ genre knowledge (see Graham et al., 1993; Graham et al., 2005). These items tend to be administered to students verbally. Student responses are divided into unique idea units, and classified according to whether they relate to production procedures, substantive processes, motivation, abilities, and/or other for writing knowledge, and production procedures, elements, and/or other for genre knowledge. A variety of writing self-efficacy measures have been employed in SRSD research. Graham et al. (1993) developed a 10-item writing self-efficacy measure; this scale has been employed in a number of SRSD studies, and has been adapted based on study contexts (e.g., Graham et al., 2005). This instrument is administered to students verbally, and students record their responses on 4-point Likert scales. The Writing Skills Self-Efficacy Scale (Pajares & Valiante, 1999) has also been used to assess students’ self-efficacy at pre- and posttest. Additionally, teachers have rated students’ intrinsic motivation and effort before and after an intervention, following procedures described by Gottfried (1990) and Maclver, Stipek, and Daniels (1991), respectively. Teachers rate both on 11-point scales, and do not have access to their pretest ratings when making posttest ratings.

Social validity of overall interventions has also been measured through interviews with teachers and students focusing on the value of SRSD strategies and the effectiveness of instructional procedures. Graham and Harris describe collecting both quantitative and qualitative data as part of their writing intervention research in the form of student and
teacher interviews and observations. The purpose of these data is to determine participants’
perceptions of the events, value, and strengths and weaknesses associated with interventions
(Graham & Harris, 2014). However, as suggested by the nature of the measures described
above, the published SRSD writing studies have a heavy quantitative focus and there are few
examples of mixed methods designs (for exceptions see Mason, Kubina, Kostewicz, Cramer,
& Datchuk, 2013; McKeown, 2012; Zumbrunn & Bruning, 2013). Thus, the SRSD research
literature could be enriched by additional mixed methods studies which incorporate a diverse
variety of measures and perspectives with the goal of further improving SRSD treatments
and better understanding the effects of SRSD interventions in terms of learners’ writing
performance, knowledge, self-regulatory strategy use, motivational beliefs, and the social
validity of interventions.

The SRSD writing model has been successfully implemented in a variety of
instructional contexts, including elementary and middle school whole-classroom instruction,
small groups, and one-on-one settings (Graham & Harris, 2003). Previous studies have
suggested that students develop proficiency with strategy use after 8-12 lessons of
approximately 30–40 minutes each; lessons are generally taught three to five days per week
(Graham & Harris, 2003; Harris et al., 2008b). Results of previous SRSD research have
suggested that struggling writers in particular do not demonstrate significant improvements
in writing performance until the final stage of instruction (Harris et al., 2008b).

To date more than 100 research studies have been conducted to examine the impacts
of SRSD writing instruction. Research examining implementation of the SRSD model in
writing has shown significant improvements in writing performance of elementary and
middle school students (Festas et al., 2015, Graham & Harris, 2003; Harris et al., 2006), high school students (Jacobson & Reid, 2012) and adults (Berry & Mason, 2012). SRSD has shown to be effective for gifted writers (Albertson & Billingsley, 1997), regularly achieving writers (De La Paz & Graham, 2002), struggling writers (Harris et al., 2006), students with learning disabilities (De La Paz, 1999), students with autism spectrum disorders (Asaro-Saddler, 2014), and students with behavioral and emotional disabilities (Ennis, Jolivette, Terry, Fredrick, & Alberto, 2015). There is some evidence that the effects of SRSD may be most pronounced for lower-performing students (Graham & Perin, 2007b). Significant gains resulting from the SRSD model of instruction have been documented in the areas of (1) writing quality, (2) writing knowledge, (4) writing self-efficacy (4) approach to writing, and (5) inclusion of genre elements. Furthermore, SRSD interventions have shown to improve students’ strategy use, including those related to planning, organizing, revising, and self-monitoring (Harris et al., 2008b). There is some evidence for the impact of SRSD instruction on students’ intrinsic motivation and effort (Harris, Graham, & Adkins, 2015), though to date these measures have been based on teacher perceptions. Researchers have also documented maintenance effects over time and found generalization effects of SRSD instruction across different settings, teachers, and writing media.

Graham, Harris, and Mason (2005) implemented SRSD interventions with SRSD, SRSD with peer support, and control conditions to investigate whether the addition of peer support would facilitate maintenance and generalization. Graham et al. (2005) found that third grade students in the SRSD and SRSD plus peer support conditions wrote longer, more complete, and higher-quality stories and persuasive essays than students in the control
condition, and the treatment effects associated with story writing continued at maintenance. The authors did not find incremental effects of the peer support component added to SRSD instruction on writing quality or knowledge about good writing and the behaviors of good and poor writers; nonetheless, they did find that peer support impacted students’ knowledge about planning and resulted in transfer effects to personal narrative and informative writing.

Brunstein and Glaser (2011) employed a path-analytic model in an effort to identify the mediators of SRSD writing strategies and writing performance. The authors assigned fourth grade students to SRSD and strategy-only conditions; the latter only differed from the SRSD condition in that it did not include self-regulatory supports. Students in the SRSD condition demonstrated greater writing knowledge, higher writing self-efficacy, and greater use of planning and strategy-related revisions, and wrote higher-quality stories than students in the strategy-only condition at both posttest and maintenance. Path analyses showed that the addition of self-regulatory supports to strategy instruction most directly improved students’ story planning, which supported students’ deep-level revisions to text and ultimately impacted the quality of their stories. Both writing knowledge and self-efficacy also predicted story planning, and thus had an indirect impact on story quality. Brunstein and Glaser’s (2011) findings illustrate how SRSD interventions influence students’ declarative and procedural knowledge in addition to their motivational beliefs and how these variables work together to impact writing performance.

Limpo and Alves (2013) examined the effectiveness of planning and sentence combining strategy instruction in the context of SRSD interventions. The authors assigned fifth and sixth grade classes to planning, sentence combining, and control conditions. Results
showed that planning and sentence combining instruction positively impacted students’ respective use of those skills and was associated with longer and higher-quality opinion essays compared to the control condition. At posttest, stronger effect sizes were seen for planning than sentence combining, and planning instruction effects generalized to summary writing while the effects of sentence combining instruction did not. Though there had been no relations between self-efficacy and writing quality for any of the conditions at pretest the authors found moderate correlations between writing self-efficacy and writing quality for both treatment conditions at posttest. Limpo and Alves (2013) hypothesized that the strategy instruction, and the focus on self-monitoring in particular, resulted in more accurate self-perceptions of students’ ability and helped to bring students’ self-efficacy beliefs more in line with their actual performance.

Ennis et al. (2015) used a piecewise hierarchal linear model (HLM) to investigate the impact of SRSD instruction for persuasive writing on the writing performance and academic engagement of middle and high school students at risk for emotional and behavioral disorders. The intervention was administered by teachers two days per week, following two-hour training and one-hour practice sessions. Results showed gains in correct word sequences, essay elements, and essay quality during the intervention compared to baseline. However, results also showed growth in all three writing performance measures was significant during weeks 1–5 of the intervention but non-significant during weeks 6–8. Students’ academic engagement growth was significant weeks 6–8 only. This study is the only current example of an SRSD study that modeled writing quality growth over time using a hierarchal model.
Finally, Hacker et al. (2015) used a quasi-experimental design to compare the effects of SRSD instruction conducted in one school to the effects of traditional writing instruction conducted in another school. Seventh grade students in both schools were preparing for a high-stakes persuasive writing assessment. Essays written online at pretest, posttest, and maintenance time points were assigned writing quality scores by the Project Essay Grade (PEG) scoring engine and growth was analyzed using HLM. Results showed no difference between the writing quality scores of students in the SRSD and comparison conditions at posttest, but students in the SRSD condition produced higher-quality essays at maintenance.

While the aforementioned studies provide evidence of the effectiveness of SRSD instruction for particular students in particular contexts, the overall effectiveness of SRSD has been verified in several meta-analyses (i.e., Graham, 2006; Graham & Harris, 2003; Graham et al., 2012; Graham & Perin, 2007a). Each of these meta-analyses examined the impacts of SRSD writing instruction on writing performance. Graham and Harris (2003) examined SRSD studies exclusively, focusing on impacts of the writing intervention model for students with learning disabilities (LD). Graham (2006), Graham and Perin (2007a), and Graham et al. (2012) examined writing intervention studies more broadly to examine impacts of instruction on writing performance. This allowed for comparing the reported effects of SRSD studies to those of non-SRSD studies; however, limitations associated with coding studies based on whether they used the SRSD model have been described in the literature. As described above, the SRSD model provides a framework for instruction and is not prescriptive; thus, implementation of SRSD instruction varies by design. Furthermore, non-SRSD studies may vary in the extent to which they incorporate elements of the SRSD model,
such as scaffolded instruction and interactive learning. Distinguishing characteristics of non-SRSD studies are that they are not criterion-based and do not emphasize SRL (Graham, 2006; Graham & Perin, 2007a).

Graham and Harris’ (2003) meta-analysis comprised the SRSD writing intervention studies conducted through 2002, all of which were conducted in grades 2–8. Based on an examination of planning strategies, the authors reported effect sizes for group designs of 1.47 for quality, 1.87 for elements, and 2.07 for length. Additionally, for single-subject designs, the average percentage of non-overlapping data (PND) points (see Scruggs & Mastropieri, 2001), was 97% for quality, 92% for elements, 100% for story grammar (all very effective treatments), and 82% for length (effective treatments). Graham and Harris (2003) also concluded that SRSD planning strategy instruction is effective for helping students with LD improve their writing, based on average effects sizes for group designs of 1.14 for quality, 1.86 for length, and 2.15 for elements, and average PNDs of 89 or higher for all variables. SRSD revising strategy instruction was found to have moderate to strong impacts on both the total number of revisions and the number of substantive revisions made by students. Furthermore, SRSD instruction had greater effects on writing quality for younger students, though effects for length were more pronounced for older students. Table 2.3 presents average effect sizes for SRSD studies at different grades, genres, and instructors.
Table 2.3

*Average Effect Sizes for SRSD Studies at Different Grades, Genres, and Instructors*

<table>
<thead>
<tr>
<th>Measures</th>
<th>Grades 2-6</th>
<th>Grades 7-8</th>
<th>Narrative</th>
<th>Expository</th>
<th>GA or researcher</th>
<th>Classroom teacher</th>
</tr>
</thead>
</table>

**Quality**
- Group-Post.    | 1.68 (4)   | 1.21 (3)   | 1.52 (3)  | 1.44 (4)   | 1.48 (5)        | 1.45 (2)         |
- Group-Maint.   | 0.82 (1)   | 0.71 (3)   | 0.82 (1)  | 0.71 (3)   | 0.73 (3)        | 0.74 (1)         |
- Group-Gen genre| -          | -          | 0.56 (1)  | 1.15 (1)   | -               | -                |

**Elements**
- Group-Post.    | 1.38 (2)   | 2.15 (2)   | 1.76 (1)  | 1.99 (3)   | -               | -                |
- Group-Maint.   | 1.16 (1)   | 1.82 (2)   | 1.16 (1)  | 1.82 (2)   | -               | -                |
- SP-Post.       | 96% (7)    | 85% (4)    | 100% (3)  | 89% (8)    | 94% (5)         | 90% (5)          |
- SP-Maint.      | 85% (7)    | 100% (4)   | 95% (3)   | 89% (8)    | 77% (5)         | 100% (6)         |
- SP-Gen genre   | -          | -          | 79% (2)   | 88% (2)    | -               | -                |
- SP-Gen P & P   | -          | -          | 100% (2)  | 100% (2)   | 100% (2)        | 100% (2)         |

**Story grammar**
- SP-Post.       | -          | -          | -         | -          | 100% (2)        | 100% (2)         |
- SP-Maint.      | -          | -          | -         | -          | 67% (1)         | 100% (2)         |

**Length**
- Group-Post.    | 2.02 (2)   | 2.10 (3)   | 2.21 (1)  | 2.04 (4)   | 2.39 (4)        | 0.82 (1)         |
- Group-Maint.   | 0.49 (1)   | 1.01 (3)   | 0.49 (1)  | 1.01 (3)   | 0.82 (3)        | 1.07 (1)         |
- Group-Gen genre| -          | -          | 0.27 (1)  | 1.58 (1)   | -               | -                |
- SP-Post.       | 71% (1)    | 85% (3)    | 71% (1)   | 85% (4)    | 71% (1)         | 85% (4)          |

*Note.* Narratives include stories or personal narratives. Expository included opinion and explanation essays. Group, group design; SP, single-participant design; Post., post-test; Maint., maintenance; Gen. genre, generalization to genre; Gen. P & P, generalization to a different person or place. Adapted from “Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies,” by S. Graham & K. R. Harris, 2003, *Handbook of learning disabilities*, p. 340. New York: Guilford Press.

Graham (2006) examined the impacts of strategy instruction in grades 1–12 on writing performance. Graham (2006) included experimental and quasi-experimental designs in a meta-analysis, as well as single-subject designs provided multiple-baseline designs had been used. Results showed an average weighted effect size of 1.57 for SRSD studies using group designs compared to 0.89 for similar non-SRSD studies. The average weighted effect
size for the SRSD studies at maintenance was 1.15; unfortunately, too few non-SRSD studies collected maintenance data to support comparisons. Single-subject SRSD designs did not differ significantly from non-SRSD single-subject studies in their effects, however. The author concluded that the effects of writing strategy instruction in general appear robust—given the overall impact seen across students, grades, genres, cognitive process and strategies taught—and the SRSD model appears to be a particularly effective writing intervention.

In an effort to determine which instructional practices improve the quality of students’ writing, Graham and Perin (2007a) conducted a meta-analysis of grade 4–12 writing intervention studies. The authors included all studies previously contained in the Graham and Harris (2003) and Graham (2006) meta-analyses, in addition to other high-quality experimental and quasi-experimental design studies, for a total of 154 effect sizes. They found the SRSD model to have a strong, positive effect on students’ writing, and reported an average weighted effect size of 1.14 for SRSD studies compared to 0.62 for non-SRSD studies. They concluded that strategy instruction in general had a strong impact on writing quality (ES=0.82), and was particularly beneficial for struggling writers and when delivered via the SRSD model. The authors called for additional research to determine why the SRSD model of writing instruction is associated with such large effect sizes.

Finally, Graham et al. (2012) conducted a similar meta-analysis to that conducted by Graham and Perin (2007a), but focused exclusively on writing interventions conducted in grades 1–6. Again, only experimental and quasi-experimental studies were considered, and again, SRSD-based strategy instruction (ES=1.17) showed a significantly larger effect on the quality of students’ writing than non-SRSD interventions (ES=0.59). The authors were also
able to test the effect of adding self-regulation instruction to strategy instruction from results of both SRSD and non-SRSD studies, and calculated a weighted effect size of 0.50. Graham et al. (2012) posited that this empirical evidence for the added benefit of teaching students how to regulate their writing strategy use helps to explain the larger effects that have been associated with SRSD interventions compared to writing instruction that does not include support for students’ SRL.

Despite the positive effects of writing interventions including those based on the SRSD model described above, researchers have called for improvements to the writing intervention research. The quality of the writing intervention research has been found to vary widely, and has criticized as being weak overall (Graham & Perin, 2007a; Graham et al., 2012; Graham & Harris, 2014). For example, Graham and Perin (2007a) point out the need for random assignment, controls for teacher effects, pretests to determine group equivalence, and measures of treatment fidelity to improve measures of writing instruction effectiveness. The writing intervention literature is replete with studies employing weak comparison conditions (Graham & Perin, 2007a), as effects may be confounded by differences in instructional time, type of writing assignment, or instructor experience between treatment and control conditions (Graham & Harris, 2014). Researchers should take steps to control for bias when evaluating responses, for example by having students use word processing software that supports students’ spelling and eliminates issues associated with handwriting differences. Graham (1999) found that surface-level feature of texts can influence rater’s judgments of writing quality. To date, writing intervention methodologies have relied heavily on the application of ANOVA-based statistics which assume independence of observations;
however, this assumption is violated when classes are assigned to conditions and students used as the unit of analysis. Graham and Harris (2014) argue that since writing interventions tend to involve students nested in classrooms and schools, multilevel models can appropriately measure individual differences and intra-individual change, and as a result “need to become ubiquitous in the writing intervention research” (p. 110).

Despite recent efforts to employ CBLEs to support students’ SRL, the SRSD intervention literature is devoid of any online applications of the writing model. This is due in part to the previously discussed challenges faced by CBLE designers in terms of scaffolding students’ learning, as well as to constraints associated with the complexity and costs of adapting or supporting the model online. Additionally, there are limitations associated with online presentation of the SRSD writing model. For example, some attributes of SRSD instruction, such as students’ active collaboration in the learning process, individualized feedback and instructional support, and criterion based instruction, present challenges in an online format. Regardless, as technology and research advances CBLEs, the field would benefit from examining how the SRSD writing model might be supported in an online format.

**Feedback**

Definitions of feedback proposed by various researchers have differed noticeably in their breadth (Knight, 2003). Hattie and Timperley (2007) defined feedback as information provided about aspects of performance or understanding, with the purpose of “reduc[ing] discrepancies between current understandings and performance and a goal” (p. 86). Sadler (1989) presented a similar, relatively narrow definition of feedback, emphasizing that
information about the aforementioned discrepancies should only be considered as feedback when used to reduce them. Conceptualizing feedback more broadly, Tunstall and Gipps (1996) presented a typology of feedback in which they framed feedback on a continuum from evaluative to descriptive. Evaluative feedback is based on norm-referenced judgments, while descriptive feedback relates to task competence. Askew and Lodge (2000) broadly yet succinctly defined feedback as all dialogue that supports learning. Finally, Butler and Winne (1995) also broadly characterized feedback as “information with which a learner can confirm, add to, overwrite, tune, or restructure information in memory, whether that information is domain knowledge, metacognitive knowledge, beliefs about self and tasks, or cognitive tactics and strategies” (p.275). Their model acknowledges the impact of feedback on internal processes such as SRL in addition to performance.

Hattie and Timperley (2007) address how discrepancies between current and desired performance are reduced in practice. Students can employ adaptive means such as trying different and more effective strategies or increasing effort; conversely, students’ use of maladaptive means such as lowering or abandoning goals also serve to reduce such discrepancies. Furthermore, teachers can select specific, appropriate goals and clarify goals for students, in addition to providing the necessary scaffolding and support to enable students to reach goals.

Hattie and Timperley’s (2007) model of feedback indicates that effective feedback provides information related to students’ (1) attainment of a goal (2) quality of performance relative to the goal, and (3) next steps. This feedback can provide information at the task, process, self-regulation, and self-levels (Hattie & Timperley, 2007), described below. Cues
inherent in feedback direct the learner’s attention to one of these levels (Butler & Winne 1995; Kluger & DeNisi, 1996). Butler and Winne (1995) argued that effective feedback provides information about cognitive activities and explicates the relations between task cues and performance by targeting knowledge and beliefs and/or cognitive processing. Feedback specificity, complexity, and length (see Shute, 2008) are factors in determining quality. For example, elaborative feedback supports better monitoring of cognitive processing and products as well as additional cognitive processing. Studies examining feedback complexity have shown inconclusive results, suggesting there are unexamined factors mediating the effect of feedback on performance (Shute, 2008).

Feedback is often offered by teachers, but can also be provided by peers, oneself, and experiences (Hattie & Timperley, 2007). Peer feedback can serve to correct, confirm, or provide suggestions related to student products or processes. Peer feedback also overcomes limitations associated with other forms of feedback in large part due to its availability. In an average-sized classroom, students can provide one another with more immediate and individualized feedback than a teacher can provide (Topping, 2009). Some studies, including examples from the limited number of studies conducted with elementary and middle school students, have shown peer feedback to be of varying quality (e.g., Hovardas, Tsivitanidou, & Zacharia, 2014). Students expressed some concerns related to the fairness of peer assessment (Sluijsmans, Dochy, & Moerkerke, 1998). However, the majority of peer feedback studies have reported adequate reliability and validity (Topping, 2009), and recent research has suggested that peer feedback can effectively substitute for teacher feedback (Gielen, Tops, Dochy, Onghena, & Smeets, 2010).
Levels of Feedback

Hattie and Timperley (2007) describe feedback as working at four levels. Task-level feedback provides information about levels of performance or understanding related to the task and is the most common form of feedback. Kluger and DeNisi (1996) propose the critical importance of task feedback is to identify misunderstandings, which requires that task feedback be specific. There appears to be an optimal level of specificity, as highly specific task feedback has been associated with poorer subsequent performance (Ganzach, 1994). Compared to more descriptive feedback, grades tend not to function effectively as task-level feedback, and have shown negative relations with performance (Black & Wiliam, 1998; Lipnevich & Smith, 2009). Task-level feedback in the form of outcome feedback has been criticized as providing the least amount of information to support SRL (Butler & Winne, 1995).

Process-level feedback provides information about the processes underlying the task, and considers the task as perceived by the student and situated in the environmental learning context. For instance, error detection strategies, in which feedback is provided by the self, may result in a shift in strategy use or help seeking and is a primary form of process-level feedback (Hattie & Timperley, 2007). Students must perceive task cues, attend to these cues, and consider the links between cues and strategy use, in order to effectively regulate learning (Butler & Winne, 1995). Process feedback has been combined with goal-setting strategies to enhance students’ task strategies and promote deeper learning (Schunk & Swartz, 1993a).

Feedback about self-regulation provides information related to planning, monitoring, and evaluating in pursuit of learning goals. The extent to which learners can benefit from this
type of feedback depends on the degree to which they are self-regulated, and thus some self-
regulatory knowledge and abilities are necessary for learners to be capable of incorporating
feedback about self-regulation. Butler and Winne (1995) argued that feedback is a catalyst in
SRL processes, given how it can activate SRL to reduce discrepancies between the current
state and the goal. The authors are mindful that though external feedback can be incidentally
or intentionally received from teachers, peers, and other sources of information, a great deal
of rich, internal feedback has the potential to be available as a result of monitoring efforts.
Students’ knowledge and beliefs influence their perception of external feedback, and how
cues are used to set and adjust goals, select strategies, and monitor subsequent progress.
“Feedback, regardless of its source, is contextualized according to a student’s prior
knowledge and beliefs before prior knowledge and tactics are applied” (Butler & Winne, p.
264).

Hattie and Timperley (2007) note that students’ abilities to self-assess and generate
internal feedback, motivation to identify and act on internal feedback, calibration related to
confidence judgments, outcome attributions, help-seeking behaviors, and self-efficacy all
mediate the effect of feedback on self-regulation. Self-assessment involves the cognitive,
motivational, and affective domains of SRL, and depends on both internal factors such as
metacognition as well as external factors such as instructional activities and assessments
(Paris & Paris, 2001). For example, self-appraisal and -management (Paris & Winograd,
1990) are aspects of metacognition that offer feedback information to support further
regulation of cognition, motivation, and behavior. Calibration accuracy can influence
motivation to seek and accept feedback, as feedback has a greater effect in situations where
confidence judgments are shown to be inaccurate and students’ goals are considered (Schutz, 1993). Feedback about self-regulation can emphasize the connections among strategies used, effort exerted, and performance to support adaptive attributions for the causes of students’ successes and failures. Additionally, directing feedback about self-regulation back to the task rather than at the self supports increased effort and self-efficacy (Kluger and DeNisi, 1996).

Finally, feedback at the self-level provides information about the individual and tends to offer little insight relative to the task. It may even divert students’ attention away from the task onto themselves. As a result, this type of feedback, most commonly found in the form of praise, offers little actionable information to students and has been shown to have minimal impact on achievement, with the exception of praise related to effort, strategy use, or other aspects germane to the task (Hattie & Timperley, 2007; Kluger & DeNisi, 1996). Hattie and Timperley (2007) indicate that due to its lack of proximity to the task, self-level feedback seldom translates to increased task understanding or engagement or results in increased self-efficacy. Additionally, feedback at the self-level has been found to mitigate positive impacts of task-level feedback when presented at the same time.

Effects of Feedback

Research examining the effects of feedback has shown impacts on both performance and motivation. Though the existence of feedback does not necessitate that it will lead to improved student outcomes (Hattie & Timperley, 2007; Shute, 2008), powerful effects of feedback have been identified through meta-analyses. From a meta-analysis of 22 studies specific to feedback in computer-based instruction, Azevedo and Bernard (1995) reported a weighted mean effect size of 0.80 associated with feedback. Kluger and DeNisi (1996), from
a review of 131 studies, reported an average effect size of 0.38 associated with feedback interventions. This overall effect was depressed by approximately one-third of studies where feedback showed negative relations with performance; this was typically when feedback at the self-level, such as praise or criticism, accompanied feedback related to the task. The authors concluded that feedback cues related to motivation or learning support performance, and that the nature of the task can moderate the relations between feedback and performance (Kluger & DeNisi, 1996). Hattie (1999), from a review of over 500 meta-analyses incorporating 180,000 studies that examined influencers of student achievement, found feedback to be one of the strongest influencers of student achievement. While the range of studies included required a broad definition of feedback, Hattie (1999) concluded that feedback was “the most powerful single moderator that enhances achievement” (p. 12). Finally, Hattie and Timperley (2007) compared the effects of feedback across 12 previous meta-analyses comprising 196 studies and nearly 7,000 effect sizes, and reported an average effect size for feedback of 0.79.

Research on the timing of feedback has been synthesized by various scholars (e.g., Azevedo & Bernard, 1995; Hattie & Timperley, 2007; Shute, 2008). Azevedo and Bernard (1995) determined that feedback is optimally presented to learners immediately rather than on a delayed basis in computer based instruction. Hattie and Timperley (2007) argue that the timing of feedback should be considered as related to the four levels of feedback, citing evidence that effects associated with the timing of feedback are moderated by the type of feedback. Shute (2008) reported inconsistent findings related to the provision of immediate and delayed feedback, positing that both types of feedback may result in both positive and
negative impacts on learning. For example, immediate feedback might support motivation but also dependence on feedback information. Con­versely, delayed feedback might support greater cognitive and metacognitive activity but not provide sufficient support for struggling learners (Shute, 2008).

There is empirical support that the effects of feedback may be enhanced by goal setting (Kluger & Denisi, 1996; Hattie & Timperley, 2007). Hattie and Timperley (2007) found that feedback was most impactful with low-complexity tasks and specific, challenging goals.

Schute (2008) is mindful that the relation between feedback and performance can be mediated by students’ emotional states. Feedback impacts students’ achievement emotions, in addition to the success expectancies and task values associated with subsequent tasks that determine student’s initial emotions for those tasks (Pekrun et al., 2007). Pekrun, Cusack, Murayama, Elliot, and Thomas (2014) found anticipated feedback to predict students’ achievement goals and achievement emotions. This finding highlights the relations among types of feedback (normative, individual improvement, or none), goals, and emotions related to a task, and offers implications for the design of classroom structures.

To summarize, effective feedback is task- rather than self-focused and provides information that students can attend to in order to reduce discrepancies between current performance and goal states (Kluger & Denisi, 1996; Hattie & Timperley, 2007). Feedback about the processing of the task and self-regulation is influential for deep processing and task mastery, while feedback about the task is influential when used to improve processing of the task and self-regulation (Hattie & Timperley, 2007). Effective feedback is presented in the
form of audio-, video-, or computer-assisted instructional feedback which relates to goals and provides students with cues or reinforcement (Hattie, 1999). Effective feedback associates performance with effort to improve engagement with and performance on academic tasks.

**Feedback in Writing**

Writing scholars have long argued that students need opportunities to practice their writing, receive meaningful feedback, and revise written work in order to grow as writers (c.f., Kellogg & Whiteford, 2009). However, as discussed above, differential impacts of feedback on students’ achievement and motivation have been noted (Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). Given the cognitive demands associated with composition, understanding how and under what conditions writing feedback is best implemented is critical to support improvements in students’ writing quality (Nelson & Schunn, 2009).

Researchers have documented positive effects of feedback on writing performance. Schunk and Swartz (1993a) reported benefits of providing fourth and fifth grade students with process goals associated with learning a writing strategy and progress feedback related to effective use of the strategy. Process goals and progress feedback were associated with writing skill, writing strategy transfer, and writing self-efficacy. The authors concluded that use of process goals emphasizes the connection between strategy use and writing quality. Similarly, providing feedback specific to their progress confirms to students that effortful application of a strategy will result in improved writing performance. Schunk and Swartz (1993a) also argued that progress feedback, which conveys to students that they are capable
of improving their writing skills, provides a “persuasive source of efficacy information” (p. 351).

In a similar study, Schunk and Swartz (1993b) assigned gifted fourth grade students to a product goal, process goal, or process goal plus progress feedback condition. At posttest, students in the process goal and process goal plus progress feedback conditions showed greater improvements in writing quality than students in the product goal condition. At maintenance, students in the process goal plus progress feedback condition showed the greatest improvements in writing quality. The authors noted that while gifted students generally monitor their work, given the complexities of writing the progress feedback may have been particularly beneficial.

An additional line of research has looked at the impact of feedback on writing fluency. Writing fluency, or students’ ability to write efficiently and accurately, has been shown to be correlated with writing quality and writing achievement (McMaster & Espin, 2007). In one study, third grade students who practiced story writing and received feedback on the length of their stories showed greater improvements in writing fluency than students who only practiced story writing (Hier & Eckert, 2014). Truckenmiller, Eckert, Codding, and Petscher (2014) conducted a randomized controlled trial to examine the impact of performance feedback on students’ writing fluency. Utilizing a multilevel model to control for school effects, retain individual differences, and accommodate missing data, the authors concluded that over a period of seven weeks performance feedback significantly improved students’ writing fluency, based on an effect size of 0.66. Performance feedback rather than writing practice was identified as the key factor in improving students’ writing fluency.
Truckenmiller et al. (2014) proposed that performance feedback might be used in conjunction with evidence-based writing practices such as SRSD interventions to more powerfully impact students’ writing achievement.

Couzijn and Rijlaarsdam (2005) examined the impact of reader observation and written feedback on the quality of ninth grade students’ instructional texts. Students were assigned to one of six conditions: treatment conditions in which, after writing an instructional text, they either (1) re-read their text and generated self-comments; (2) observed a reader following their instructions; (3) observed a reader following their instructions and received written comments from the reader; (4) observed a reader following a peer’s instructions; (5) performed a set of instructions and produced oral feedback for the writer; (6) performed a set of instructions and produced oral and written feedback for the writer; or a control condition in which they only wrote an instructional text. Results showed a significant effect of feedback on revision quality for all observation feedback conditions, with the greatest effect being seen for students who both observed a reader following their instructions and received written comments from the reader. For the students who wrote feedback, following a set of instructions and producing oral and written feedback for the writer was also associated with higher revision quality than conditions that involved just writing or writing and generating self-comments related to an instructional text. Results supported the idea that writers are blind to some flaws in their own texts. Authentic readers’ responses to texts can function as feedback which emphasizes readers’ needs and suggests ways to revise texts to support improvements in quality (Couzijn & Rijlaarsdam, 2005).
In contrast to the majority of feedback studies which have utilized quantitative methods, Hawe and Dixon (2014) employed a qualitative case study analysis of teachers’ beliefs and knowledge about feedback and perceptions of practice. They argue for the need for students to develop evaluative capacity in writing to reduce dependency on externally-provided feedback and support development of internally-generated feedback. The authors call for increased authentic opportunities for students to evaluate, self-monitor, and revise their writing.

Despite reviews showing teacher feedback in writing contexts to be preferential to no feedback (Biber, Nekrasova, & Horn, 2011), researchers have cited concerns with feedback provided by teachers. In an examination of the nature of written feedback to the writing of third and seventh grade students, Clare, Valdes, and Patthey-Chavez (2000) found teacher feedback focused on surface-level text features. Feedback tended not to address content or organization of papers, and as students revised texts their improvements were isolated to the area of mechanics. Matsumura, Patthey-Chavez, Valdez, and Garnier (2002) similarly found third grade teachers generally focusing feedback on surface features of writing rather than effective communication of information. The authors found a negative relationship between the quality of teachers’ goals and the quality of conventions in students’ writing.

A limited body of research exists examining peer feedback, particularly in the elementary and middle grades. Olson (1990) concluded that peer feedback had a positive impact on sixth grade students’ writing quality though effects were isolated to initial drafts of stories. Students who received revision instruction and revised individually made more revisions than students who received the same instruction but revised with a peer, though
these revisions did not result in higher-quality stories. Gielen et al. (2010) found no difference in performance between middle school students who received peer feedback on assignments and those who received teacher feedback. The researchers found the greatest gains in posttest performance associated with an extended feedback condition in which writers also completed a question form to direct the peer reviewer’s attention to those aspects of writing on which the writer focused or found most difficult. Nicolaidou (2013) described benefits of using e-portfolios to document fourth grade students’ writing growth and support the use of peer feedback. Impacts of a year-long intervention were seen in improvements in students’ writing performance and their ability to provide corrective feedback to peers. Qualitative data showed that students valued their peers’ feedback, and that peer feedback had greater benefits for average and high-ability students who needed less support than low-ability students in providing feedback and incorporating feedback when revising.

Finally, Graham and Perin (2007a) have criticized the overall scarcity of research examining feedback in the context of written work and argued that more empirical evidence is needed to understand the impact of feedback on the quality of students’ writing. Graham and Perin’s (2007a) meta-analysis of writing instruction included five effect sizes of feedback on writing quality, which ranged from −0.65 to 2.52. The authors concluded that writing feedback research has varied methodologically and shown inconsistent outcomes. The vast majority of the writing feedback research has examined higher education contexts with undergraduate and graduate student participants. Also, this research has almost exclusively involved native English speaking students and English as a Second Language Learners. Researchers have expressed concerns about the lack of ecological validity of the
writing feedback research, and thus the pedagogical relevance (Lee, 2014). The writing feedback literature can be enriched by future research that examines writing feedback in authentic elementary and middle school contexts, involves a greater diversity of participants, and incorporates both quantitative and qualitative methods.

**Automated Feedback**

Sustained deliberate practice is needed to develop expertise in writing (Kellogg & Whiteford, 2009). This requires effortful activity to improve performance; intrinsic motivation for task engagement; frequent practice opportunities; feedback based on performance; and extensive practice (Ericsson, 2006). Previous research has suggested that students will reap the greatest benefit from feedback that is immediate, specific, detailed, and addresses both surface-level and content features of writing (Patthey-Chavez, Matsumura, & Valdes, 2004).

However, researchers have documented that teachers often do not provide students with writing assignments that are of a high quality (Matsumura et al., 2002). Additionally, grading of essays and the provision of high-quality feedback to student writing is time consuming for teachers (Dikli, 2010). Teachers have been found to limit writing assignments in an attempt to make the time and effort associated with grading such assignments manageable (Kellogg & Whiteford, 2009). Furthermore, educator grading has been shown to be unreliable (Harlen, 2005).

The implications of these issues are (1) students do not tend to receive feedback in a timely manner conducive to reflection, revision, and ultimately learning, and (2) students may receive mixed messages about their performance and those facets of their writing
needing improvement as a result of potentially unreliable—and thus, invalid—evaluations of
and feedback based on their writing. Low-quality evaluations of performance and feedback
are poor indicators to students of their writing abilities, and have negative implications for
the development of students’ writing skills and the motivational aspects related to their
writing.

The use of automated essay scoring (AES) and automated feedback provides more
timely feedback to students than could a teacher or peer reviewer and overcomes some of the
limitations associated with the hand-scoring of essays. AES systems are capable of providing
students with immediate evaluations of essay quality. This can be in the form of a holistic
score, trait scores, or a combination of both. Previous studies have shown AES systems to be
as or more reliable than multiple human raters in assigning scores. Shermis, Garvan, and
Diao (2008) report exact and adjacent rates achieved by AES systems with the scores of
human raters in the mid-80s and mid-high 90s, respectively, which slightly exceeds the
agreement rates achieved by experienced human scorers. Moreover, AES systems are
perfectly reliable in ways human raters are not (Shermis & Hamner, 2013). AES further
avoids the time and cost commitment associated with training humans to score essays
reliably.

In general, AES systems employ statistical models based on human evaluations of
writing in order to mimic the scoring of humans in assigning scores and providing feedback
to essays. Using natural language processing techniques such as Latent Semantic Analysis
(Landauer, Foltz, & Laham, 1998), AES systems parse and examine human-scored “training”
essays for hundreds of variables reflecting intrinsic characteristics of writing quality, build
statistical models based on findings, and then apply these models to predict holistic or trait scores for subsequent essays. In this manner, the score(s) an AES system assigns to an essay is based on the scores human raters assigned to similar essays on which the AES system was trained. Thus, the score(s) assigned by an AES system reflects the characteristics of writing shown to be valued by humans based on the training essays.

Recently, AES systems have been combined with automated feedback in automated writing evaluation (AWE) programs. Commercially available examples include Criterion from the Educational Testing Service using the e-rater scoring engine, MyAccess! from Vantage Learning using the IntelliMetric scoring system, PEG Writing and NC Write from Measurement Incorporated using the Project Essay Grade (PEG) scoring engine, and WriteToLearn from Pearson Education using the Intelligent Essay Assessor scoring engine. As AWE programs are intended to support writing instruction, they commonly include some form of learning management system to organize students’ writing portfolios and track progress, as well as an online writing lab that includes writing supports such as rubrics and graphic organizers (Grimes & Warschauer, 2010).

Writing Pal (W-Pal), which uses the Coh-Metrix computational tool, is a CBLE which has been described as an Intelligent Tutoring System (ITS, Roscoe & McNamara, 2013) that emphasizes strategy instruction, extended practice, and formative feedback. Strategy instruction is provided via 15–30 minute videos that use animated agents to explain and model writing strategies. Strategy instruction addresses prewriting (planning and organizing), drafting (writing an introduction, providing support, writing a conclusion), and revising (editing text for cohesion and clarity). Checks for understanding and quizzes are
included as part of the lessons. W-Pal also includes game- and essay-based practice, the latter supported by automated scoring and feedback. To align with the emphasis on strategy instruction and formative feedback, W-Pal does not provide feedback related to spelling, mechanics, grammar, or style (Roscoe & McNamara, 2013).

Despite the prevalence of positive contributions of AES for assessment and instruction (Shermis et al., 2008), a number of limitations have been associated with the use of AES and automated feedback. AES systems are dependent on the quality of the (human-scored) training data, and these data place a ceiling on the obtainable quality scores assigned by AES systems. If there is not clear consensus among humans about what constitutes high-, medium-, and low-quality writing for a given writing purpose or essay, there will be similar noise in the scoring models. Human scorers have been found to consider a number and variety of combinations of linguistic features when evaluating essays, regardless of scoring rubric (Crossley, Roscoe, & McNamara, 2014).

Deane (2013) describes three primary issues raised by opponents of AES. One objection relates to the appropriateness of measurement and scoring in light of the construct of writing. AES systems have been faulted for measuring a restricted writing construct, in particular for assessing text quality rather than writing ability. A second objection relates to the implications of machine-scoring on the behavior of writers. Critics have proposed that in high-stakes testing situations in particular, knowing that an essay is being machine-scored may result in a writer modifying his/her behavior in a manner that compromises the construct being assessed. A final objection relates to the implications of the technical limitations of current AES measurement methods in terms of overall trust in AES systems. Deane (2013)
concludes that the general concerns related to AES “involve a tension between ease of measurement and construct coverage” (p. 20).

**Impact of automated feedback**

Previous research examining the implications of providing students with automated feedback has shown mixed results. A number of studies report a positive effect of automated feedback on the writing quality of K-12 students (Franzke, Kintsch, Caccamise, Johnson, & Dooley, 2005; Shermis et al., 2008; Wade-Stein & Kintsch, 2004). Benefits have been found in terms of time on task and the content of students’ writing. In a study involving the use of Criterion by 6–8 and 10th grade students, Shermis et al. (2008) reported gains in overall essay quality across seven prompts, though the methodology of this study warrants cautious interpretation of results (see Wilson et al., 2014). There is some evidence that automated feedback does not have a deleterious effect on the amount or type of feedback provided by teachers, but is instead associated with proportionally more teacher feedback focused on higher-level writing skills such as development of ideas (Wilson & Czek, 2016).

Despite the aforementioned benefits of automated feedback for students’ writing, researchers have had difficulty demonstrating effects of automated feedback on overall writing quality. In an exception, Wilson et al. (2014) examined data from grade 4–8 students who used PEG to participate in statewide benchmark writing assessments. The authors used HLM to investigate the relations between essay revision based on automated feedback and improvements in essay quality. Results showed a positive impact of automated feedback on writing quality, with each revision associated with an average increase in overall score of 0.5. Differential effects of automated feedback on students were observed, suggesting that some
student characteristics may mediate the impact of feedback on writing quality. No transfer effects across prompts were found.

Finally, from a meta-analysis of the effects of word processing programs on the writing quality of weaker writers, Morphy and Graham (2012) reported an average effect size of 1.46 associated with programs providing text quality feedback or prompting planning, drafting, or revising. While only three programs were included in the analysis and they differed in complexity, these showed particularly powerful effects on writing quality.

Some authors have been more skeptical about the current ability of automated feedback to improve students’ writing. Wilson and Andrada (2013) found automated feedback insufficient for improving the overall quality of struggling writers’ essays, concluding that supports to improve the writing knowledge of struggling writers need also be in place in order to impact students’ writing performance. From a critical review of AWE studies, Stevenson and Phakiti (2014) reported modest evidence of a positive effect of automated feedback on writing quality, and inconsistent findings related to the impact of automated feedback on writing proficiency more generally. The authors cited numerous methodological flaws with the extant AWE feedback research and called for more robust methodologies and a greater research focus on effective integration of AWE programs in writing classrooms.

A small number of qualitative and mixed-methods studies have been conducted to begin to investigate how users experience and understand AES, automated feedback, and related technologies. In an examination of students’ conceptions of feedback, Calvo and Ellis (2010) found that students considered the purposes of both automated and human feedback to
be the same (i.e., either cohesive or fragmented). The “deep” learners in their study saw writing feedback as an opportunity to learn more about the topic they were writing about, while the “shallow” learners associated feedback with improvements to surface features of the text and an opportunity to improve their grades. Despite these perceptions, deep learners earned higher grades on the final writing assignment. Calvo and Ellis (2010) recommend that the purpose of feedback in general—and automated feedback in particular—be explicated for students. Though this study was conducted with undergraduate students, the findings are similar to those of studies involving K-12 students that have suggested automated feedback is largely used to make surface-level revisions (Warschauer & Grimes, 2008; Wilson & Andrada, 2013). This may be due to students using a more restricted, surface-level focused set of evaluative criteria than teachers (Varner, Roscoe, & McNamara, 2013).

Finally, researchers have discussed the challenges associated with the provision of automated feedback. Azevedo et al. (2010) acknowledged the difficulty of providing elaborative feedback in CBLE’s due to the need to have an accurate student model. Furthermore, Wilson et al. (2014) raise questions related to whether students would maximally benefit from being provided with additional types of feedback such as those relating to strategies and process in addition to performance, as well as whether students require additional instruction to be able to fully benefit from automated feedback in terms of using the feedback to improve writing in meaningful ways rather than just improving surface features of the text.
Impact of AWEs and related technologies

The focus of AES research to date has been the psychometric properties of scores assigned by AES systems. This emphasis on the assessment- and psychometric-related aspects of AES has resulted in an insufficient understanding of classroom applications and pedagogical implications of AWE programs in their ability to improve the quality of students’ writing (Stevenson & Phakiti, 2014). Little research has examined the impacts of AWE programs on the quality of K-12 students’ writing. Findings have generally shown positive effects of AWE on a range of smaller-grain writing outcomes. These programs have been demonstrated to be successful in improving the mechanical aspects of writing and have been associated with increased essay length, though there is a lack of evidence that they are powerful enough to improve overall writing quality (Rock, 2007; Roscoe & McNamara, 2013; Shermis, Burstein, & Bliss, 2004). Wilson and Andrada (2013) expanded on such findings by analyzing students’ patterns of performance within AWE programs. Their findings suggested that struggling writers do not use automated feedback to make substantive improvements to essays, but rather use feedback to edit their compositions for spelling and grammar. Some researchers have suggested that students tend to make an insufficient number of revisions in AWE programs, submitting the majority of essays a single time (e.g., Attali, 2004; Warschauer & Grimes, 2008).

Warschauer and Grimes (2008) collected a variety of qualitative data to examine classroom use of MyAccess! and Criterion in grades 6–12. Teachers reported time-saving benefits of the programs in terms of reducing grading and supporting individualized instruction. However, there was not a commensurate increase in student writing opportunities,
in part due to an insufficiency of relevant writing prompts. Results also showed that while students generally implemented the feedback they received, revisions largely addressed surface-level aspects of compositions. The authors also noted that MyAccess! and Criterion did not appear to negatively influence how writing was taught, nor to encourage formulaic writing, as has been a concern of critics.

Grimes and Warschauer (2010) conducted a similar study to examine attitudes towards and instructional practices with MyAccess! across eight middle schools. Teachers again reported generally positive benefits of MyAccess! for classroom management, as it was seen to support student autonomy and motivation for writing and assist with writing portfolio management. Grimes and Warschauer (2010) found no evidence of student attempts to game the software despite teachers reports to the contrary. The authors posited that gaming attempts in AWE programs are not of concern, as these likely prompt students’ critical analysis of the differences between human and machine scorers. Both teachers and students reported increased amounts of revision using MyAccess!. The feedback was criticized as being overly extensive to the point of being overwhelming for students. Students were found to incorporate surface-level revisions prior to attempting to address structural issues. The authors also cited concerns that both teachers and students placed undue faith in MyAccess! in terms of placing unwarranted trust in trait scores and assuming that the software made human-like judgments in evaluating essays.

An initial trial of W-Pal over six months which focused on tenth grade students’ perceptions of the utility and design of the program showed W-Pal to be a generally useful system of writing instruction, practice, and feedback (Roscoe & McNamara, 2013). Posttest
essays written outside of W-Pal showed improvements in the structure and lexical sophistication as well as the overall quality of students’ writing; however, methodological limitations did not support attributing these effects to W-Pal. Students expressed concerns that the strategy instruction lessons were too long and insufficiently engaging, and also criticized the lack of feedback specificity, which was seen to undermine their overall confidence in the program.

In a follow-up study, Proske, Roscoe, and McNamara (2014) found that students assigned to a game-based practice condition found W-Pal more engaging than students assigned to question-based practice, and equally engaging as students assigned to model- and writing-based practice conditions. Nonetheless, students in the model- and question-based practice conditions showed the greatest application of writing strategies based on a 10-item, multiple choice posttest. The authors hypothesized that by encouraging reflection, the model-based practice supported students’ learning. Scaffolding and feedback were inferred to be necessary elements of the ITS to optimally support students’ application of writing strategies.

Roscoe and McNamara (2013) note the need to consider the effectiveness of technologies that grade essays, present summarization and argumentation instruction, and provide scaffolds for students’ writing for more broadly addressing the pedagogical needs associated with the ill-defined nature of writing. Given the assessment focus of current AWE programs, practice and feedback are emphasized to the detriment of writing strategy instruction. Without the support of strategy instruction students may be incapable of implementing the feedback they receive (Roscoe & McNamara, 2013). There is additionally a need for further research that examines both teachers’ and students’ perceptions of AWEs
as CBLEs (Grimes & Warschauer, 2010). The nascent research examining classroom applications of AWEs is promising despite effectiveness being moderated by student and teacher perceptions of automated scores or feedback (Roscoe & McNamara, 2013).

**Overview of the Present Study**

SRL is critical in writing contexts given the importance of writing well and the inherent complexity of writing. Over 30 years of research has validated SRSD as a uniquely powerful model of writing instruction. This is likely because it addresses both the cognitive and self-regulatory skills needed to write effectively. As the focus of AWE research to date has been on the validity of these programs (i.e., comparability to human scoring), much remains to learn about practical applications of AWE programs. A writing intervention that combines SRSD with an AWE program addresses significant gaps in both the SRSD and AWE literatures. First, such an intervention provides an opportunity to examine how the SRSD model might be adapted to take advantage of AWE technology while examining the efficacy of an intervention designed to be implemented by practitioners. The vast majority of SRSD interventions described in the literature have been implemented by researchers or tutors in small group or one-on-one settings rather than classroom teachers. Second, there is a need to examine strategy instruction in order to support students’ use of AWE programs (Roscoe & McNamara, 2013). Third, teachers’ and students’ perceptions of AWEs would benefit from a close examination to better understand the implications of classroom application (Grimes & Warschauer, 2010; Roscoe & McNamara, 2013). Finally, the SRSD research has almost entirely been conducted by a single research group, thus there is value in
providing external validation and evidence of scalability for the SRSD model (Baker, Chard, Ketterlin-Geller, Apichatabutra, Doabler, 2009; Hacker et al., 2015).

The present study addressed the four aforementioned gaps in SRSD and AWE literatures in the following ways. First, the present study utilized an intervention design that leveraged the strengths of the SRSD model with the writing practice opportunities and immediate feedback afforded by an AWE program. The outcome was a treatment that could be administered by teachers a couple days a week without extensive training using the AWE program NC Write. Second, lessons based on validated approaches to strategy instruction were developed for NC Write in response to the need for AWEs to shift from evaluating writing to supporting writers (Roscoe & McNamara, 2013). Strategy instruction was an integral part of lessons in both the classroom and in NC Write. Additionally, practice and feedback in NC Write were incorporated as tools, not ends in themselves. The use of two treatment conditions allowed for differentiating SRSD support from traditional writing instruction support. Third, qualitative data collection and analysis supported an examination of teachers’ and students’ perceptions of NC Write. Finally, the design and implementation of this study was conducted independent of the research team responsible for the majority of the extant SRSD research.

The present study investigated the effects of NC Write + SRSD instruction and NC Write + traditional writing instruction on students’ writing quality, knowledge, and self-efficacy. Students in treatment conditions received SRSD instruction or traditional writing instruction from their classroom teacher. These students also used NC Write to complete
instructional lessons, practice writing essays, and receive scores and feedback based on their writing.

SRSD instruction provided by teachers and reinforced in NC Write comprised a general planning strategy, argumentative-specific strategies, and the knowledge and self-regulatory strategies necessary for implementing the writing strategies successfully based on Graham and Harris’s (1993) SRSD model. A key aspect of this model is that students “learn how to use the self-regulation procedures, including goal setting, self-monitoring, self-reinforcement, and self-instructions, to help them manage the writing strategies and task of writing as well as to obtain concrete and visible evidence of their progress” (Harris et al., 2006, p. 306).

Traditional writing instruction provided by teachers and reinforced in NC Write addressed the basic structure of argumentative essays, planning and organizing strategies, and a variety of relevant writing skills.

Sixth through eighth grade students were assigned by teacher to one of three conditions: NC Write + SRSD instruction (NC + SRSD), NC Write + traditional writing instruction (NC + TRAD), or a comparison condition (COMP) that received traditional writing instruction and did not use NC Write. An embedded quasi-experimental after-intervention (QUAN + (qual)) mixed methods design (Creswell & Clark, 2008) was used. The intent of this design is to embed qualitative data collection within a quantitative experiment to supplement the quantitative results (Creswell, 2015). Appendix A presents an overview of the research design. The present study addressed three research questions:
1. What changes in students’ writing performance, knowledge, and self-efficacy are associated with NC Write + traditional writing instruction and NC Write + SRSD instruction compared to traditional writing instruction?

It was predicted that students in the NC + SRSD condition would display greater gains in writing quality and knowledge than those in the NC + TRAD and COMP conditions based on previous research comparing the effects of SRSD interventions to other writing interventions (Graham & Perin 2007a; Graham et al., 2012). SRSD instruction was not predicted to be associated with greater gains in self-efficacy than the other two conditions. Previous research has found inconsistent impacts of SRSD instruction on self-efficacy. Though the SRSD model is designed to support students’ self-efficacy, their efficacy beliefs may drop as a result of greater self-regulation, in particular more accurate self-perceptions of ability (Limpo & Alves, 2013). It was also predicted that females would demonstrate greater gains in writing performance than males. Prior research has documented gender differences in writing development, and authors have reported females producing both longer and higher-quality compositions than males (Berninger & Swanson, 1994, Olinghouse, 2008).

2. What is the shape of growth in students’ writing quality, essay length, and essay elements associated with NC Write + traditional writing instruction and NC Write + SRSD instruction?

Students in the treatment conditions were predicted to show growth in writing quality, essay length, and essay elements over the course of the intervention. It was unknown whether this growth would be at a constant rate over the course of treatment or take on a curvilinear shape. Previous SRSD research has shown that struggling writers do not demonstrate
significant improvements in writing performance until the final stage of instruction (Harris et al., 2008b), suggesting accelerated growth towards the end of treatment. However, Wilson et al. (2014) found writing quality growth based on essay revisions using automated feedback to be nonlinear and concave in shape (Wilson et al., 2014). Thus, both linear and quadratic growth was examined.

3. What are students’ and teachers’ perceptions of the events, value, and strengths and weaknesses associated with NC Write?

Previous research has described the moderating effects of student and teacher perceptions of automated scores or feedback on writing quality (Roscoe & McNamara, 2013). Additionally, researchers have noted that further qualitative and mixed-methods studies are needed to better understand adoption practices of technological writing interventions as well as teachers’ and students’ perceptions of these interventions. Given that NC Write employs many technologies that are new to users, it was important to investigate how users perceived the program as a step towards understanding those aspects of the program that hinder and support improvements in students’ writing quality.
CHAPTER THREE

Methods

Participants

In the summer of 2015, school districts across North Carolina that were not current or former users of NC Write were contacted to gauge interest in participation in the study (see Appendix B). Districts were targeted to provide racial and socioeconomic diversity to the sample and to include participants from the coastal plain (eastern), piedmont (central), and mountain (western) regions of the state. As an incentive to participate districts were offered free NC Write accounts for participating teachers and their students. Five districts expressed interest in participation. Once these districts provided information regarding existing structures, schedules, and flexibility, teachers were recruited within districts to participate in one of the three conditions. For example, some schools or grade-level teams of teachers within schools used established writing programs which made them suitable candidates for the COMP condition. Other schools or teams of teachers had greater flexibility in administering an intervention making them suitable for one of the treatment conditions. Multiple conditions were represented within both districts and schools. A total of 19 teachers participated in the study. None of the teachers were aware of the predicted outcomes of the study.

Students were recruited from participating teachers’ classrooms. All students in participating classrooms had access to NC Write and the study materials but only the data of those students who provided informed consented to participate in the study were analyzed. At
the start of recruitment 1404 students were enrolled in participating teachers’ classrooms. Informed consent was ultimately provided by 978 of these students. Of the consenting students, 149 were excluded from the sample for one or more of the following reasons: not completing the pretest essay; not completing the posttest essay; completing fewer than three essays total during the intervention period; and/or changing class/teacher assignment during the intervention period. The final sample comprised 829 students in grades 6 through 8. Participant demographics are presented in Table 3.1.
Table 3.1. Participant Demographics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Comparison</th>
<th>NC Write + traditional writing instruction</th>
<th>NC Write + SRSD instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students (n)</td>
<td>270</td>
<td>272</td>
<td>287</td>
</tr>
<tr>
<td>Districts (n)</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Schools (n)</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Teachers (n)</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Grade (n)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>57</td>
<td>41</td>
<td>113</td>
</tr>
<tr>
<td>7</td>
<td>128</td>
<td>160</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>85</td>
<td>71</td>
<td>146</td>
</tr>
<tr>
<td>Male (%)</td>
<td>49.06</td>
<td>48.71</td>
<td>45.10</td>
</tr>
<tr>
<td>Race (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>61.29</td>
<td>78.71</td>
<td>50.74</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>13.98</td>
<td>12.93</td>
<td>12.59</td>
</tr>
<tr>
<td>Black or African American</td>
<td>19.89</td>
<td>4.18</td>
<td>32.96</td>
</tr>
<tr>
<td>Asian</td>
<td>1.08</td>
<td>1.14</td>
<td>37</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>-</td>
<td>.38</td>
<td>.37</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>-</td>
<td>.38</td>
<td>-</td>
</tr>
<tr>
<td>Two or More Races</td>
<td>3.76</td>
<td>2.28</td>
<td>2.96</td>
</tr>
<tr>
<td>Free or Reduced Lunch (%)</td>
<td>49.00</td>
<td>60.33</td>
<td>59.96</td>
</tr>
<tr>
<td>Limited English Proficiency (%)</td>
<td>2.34</td>
<td>3.06</td>
<td>5.71</td>
</tr>
<tr>
<td>Students with Disabilities (%)</td>
<td>6.87</td>
<td>5.68</td>
<td>14.12</td>
</tr>
<tr>
<td>Age (M, SD)</td>
<td>156.54 (9.93)</td>
<td>155.51 (9.11)</td>
<td>156.42 (12.31)</td>
</tr>
<tr>
<td>Previous ELA achievementc (M, SD)</td>
<td>452.70 (11.27)</td>
<td>455.67 (10.53)</td>
<td>454.53 (11.61)</td>
</tr>
<tr>
<td>Achievement goal orientation (M, SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery approach</td>
<td>12.63 (2.41)</td>
<td>12.86 (2.30)</td>
<td>12.97 (2.27)</td>
</tr>
<tr>
<td>Performance approach</td>
<td>10.03 (3.38)</td>
<td>9.49 (3.24)</td>
<td>10.21 (3.25)</td>
</tr>
<tr>
<td>Mastery avoid</td>
<td>8.22 (3.51)</td>
<td>7.51 (3.23)</td>
<td>7.99 (3.26)</td>
</tr>
<tr>
<td>Performance avoid</td>
<td>11.84 (2.96)</td>
<td>11.77 (2.94)</td>
<td>11.93 (2.98)</td>
</tr>
</tbody>
</table>

Note: N = 829.

*a*Free or Reduced Lunch data were not available for individual students and were estimated based on school-level data.

*b*Age in months at time of pretest

*c*Previous ELA Achievement based on scale score (range: 423–484) from the previous year’s end-of-grade assessment.

Consenting students were further sampled to participate in interviews. A total of 30 students were interviewed, 15 per treatment condition. Students in the interview sample represented four districts, five schools, and ten different teachers. Additionally, 12 teachers,
six per treatment condition, were interviewed. These teachers represented five districts and six different schools.

**Study Context: NC Write**

The AWE program NC Write includes interactive student lessons, writing prompts, electronic graphic organizers, automated scores and feedback, and writing portfolios. NC Write utilizes the PEG engine to provide students with automated scores and feedback for their essays. PEG statistically analyzes essays, calculates measures reflecting the intrinsic characteristics of writing, and models the decisions of professional raters in producing scores. Specifically, PEG uses a syntactic parser to measure millions of text features which are included in a variety of algorithms that correlate the features with human scores. Feature identification involves a handcrafted feature set of elements of text that commonly correlate with high and low scoring responses which is extended with a theoretically infinite set of intrinsic features that can be extracted from the training set. These may include involve n-grams of characters, words, parts of speech, and phrases; similarly measures; and matrix manipulations on the underlying intrinsic and extrinsic features. If these features correlate with scores given by professional raters they are encoded into the model regardless of whether they are directly observable. PEG’s modularity allows for algorithms and techniques to be swapped in and out to keep up with current automated scoring technology. PEG is perfectly reliable and free from error associated with multiple human scorers. In Phases one and two of the Automated Student Assessment Prize (ASAP) competition sponsored by the Hewlett Foundation, PEG achieved the highest level of agreement of all competitors.
measured against the human scores, and was found to be more reliable than two professional scorers (Morgan, Shermis, Van Deventer, & Vander Ark, 2013; Shermis & Hamner, 2013).

PEG is currently used in large scale formative and summative assessments across the country. For example, in 2013-2014 PEG was used by the Smarter Balanced Assessment Consortium to score nearly three million ELA essays and ELA and math short answer responses for the pilot and field tests. A number of Smarter Balanced Assessment Consortium member states, including CA, MI, SD, VT, and WI, have utilized PEG for operational scoring between 2014-2015 and the present, during which time PEG has scored tens of millions of student responses. PEG has also provided scores for tens of millions of formative assessments for students in grades 3-12.

PEG typically evaluates narrative, argumentative, and informational essays on six writing traits: development of ideas, organization, style, language, sentence structure, and conventions. Rather than developing prompt-specific scoring models, PEG’s training data for argumentative writing included a wide variety of argumentative essays and the resultant model was capable of accurately scoring responses to any argumentative writing prompt. For each essay submitted, students receive a score report that includes a marked-up version of their essay annotated with spelling and grammar feedback (Figure 3.1). The score report also includes a total essay score (out of 30) and trait-specific scores (each out of six) based on development of ideas, organization, style, word choice, sentence structure, and conventions (Figure 3.2). Finally, the score report includes a writing analysis in the form of evaluation and feedback for each trait (Figure 3.3).
Figure 3.1. Student essay annotated with spelling and grammar feedback.

Figure 3.2. Essay total and trait-specific scores.
Each interactive lesson in NC Write is associated with one of the six writing traits noted above, and classified as being of beginner, intermediate, or advanced difficulty (Figure 3.4). Lessons are no more than 20 minutes in length. For accessibility purposes most lessons include read-aloud functionality, where a student may hear any portion of text by hovering over a speaker icon. Lessons include a combination of animation, audio, and interactive checks for understanding to engage students. Lessons incorporate a range of 16 possible interactions, such as summary (Figure 3.5) and drop-down (Figure 3.6).
Figure 3.4. Lessons in *NC Write*.
Figure 3.5. Summary lesson interaction.
For the present study, students in the treatment conditions received SRSD or traditional writing instruction from their classroom teacher while using NC Write to complete interactive lessons, write essays, and receive scores and feedback to guide their writing. During the course of the intervention students in these conditions completed six interactive lessons in NC Write, wrote five essays (excluding pretest/posttest essays), received scores and feedback for each essay, and had the opportunity to revise each essay using the provided feedback.
Time Frame and Conditions

The study was conducted during the 2015-2016 school year. One teacher in the SRSD condition piloted the intervention in the fall of 2015; all other teachers administered the intervention in the spring of 2016. Students completed pretest measures over a period of one week. The intervention lasted from eight to twelve weeks depending on how teachers incorporated it into their schedules. On average, the intervention was conducted approximately twice per week for 30–45 minutes per class period. Students completed the posttest measures the week following the intervention.

Students in the NC + SRSD instruction condition received argumentative writing instruction based on the SRSD model from their classroom teacher. These students used NC Write to access interactive lessons that emphasized self-regulatory strategies, write argumentative essays, and revise essays using feedback provided by the program.

Students in the NC + TRAD condition received traditional argumentative writing instruction from their classroom teacher. Traditional writing instruction addressed the basic structure of argumentative essays, planning and organizing strategies, and a variety of relevant writing skills. These students used NC Write to access interactive lessons that focus on the development of ideas, organization, word choice, sentence structure, and conventions; wrote argumentative essays; and revised essays using feedback provided by the program.

A third condition served as a comparison condition; students assigned to this condition received traditional writing instruction as described above from their classroom teacher and did not use NC Write other than to complete pretest and posttest essays.
Instruments and Measures

**Writing performance.**

**Writing prompts.** Writing assessments involved writing an argumentative essay in response to a writing prompt. At pretest and posttest, prompts were counterbalanced across time points. A total of three prompts were counterbalanced as the original design included assessing students at a maintenance time point; unfortunately, posttest measures were completed too late in the school year for this to occur. During the intervention, students selected from a choice of two prompts for each writing assessment.

All writing prompts were reviewed by content-area specialists for appropriateness and interest. Prompts addressed topical issues, such as those involving school and public policy impacting middle school students. One pair of prompts was source-based and required that students read a series of brief articles and watch a video prior to composing their responses. For each writing assessment students wrote about a single topic in response to the assigned or chosen prompt.

Several steps were taken to ensure equivalency of the pretest/posttest prompts. First, these prompts were selected from a bank of prompts that had been administered operationally on a statewide writing assessment in the Northeastern United States. Previous operational use showed these three prompts to be of comparable difficulty ($M = 6.86, 6.95, \text{ and } 6.96$ on a 12-point scale). Additionally, one-way ANOVAs were conducted to examine for significant differences in pretest essay quality, length, and number of elements across the three prompts. Results showed no significant differences in pretest essay quality $F(2, 826) = .57, p = .57,$
length $F(2, 826) = 1.24, p = .29$, or number of elements $F(2, 826) = .11, p = .90$ across the three prompts.

Teachers were provided explicit directions for administering pretests and posttests (see Appendix C). Students had as much time as needed to write their essays. For the pre- and posttests, revising was disabled in NC Write and students were only able to submit a single draft. During the intervention, students had the opportunity to revise each essay submitted. Only the last draft submitted for each prompt was analyzed. The first essay assigned during the intervention was written collaboratively and was therefore excluded from analysis. All writing prompts appear in Appendix D.

**Writing quality.** PEG provides measures of writing performance for each of six traits: (1) *Development of ideas*, based on use of reasons, details, and examples; (2) *Organization*, based on control; (3) *Style*, based on demonstrated awareness of audience, task, and writing purpose; (4) *Word choice*, based on use of vocabulary; (5) *Sentence structure*, based on usage and variety of sentences, and (6) *Conventions*, based on use of spelling, punctuation, and capitalization. For each trait, quality is measured on a 1–5 point scale. A sample rubric appears in Appendix E. In the present study the PEG total essay score was used as the outcome variable as individual trait scores have been found to be highly correlated (Wilson et al., 2014).

**Essay length.** The length of students’ essays was measured in number of words. This measure included all written words regardless of spelling. Essay length was calculated using Microsoft Excel.
**Essay elements.** Essays were evaluated for the presence of the basic elements of argumentative writing (Scardamalia, Bereiter, & Goelman, 1982), including claim, supporting reasons, elaborations, counterclaims, and conclusion. For both claim and conclusion, a score of 1 was awarded if the element was present and 0 if it was not present. For supporting reasons, elaborations, and counterclaims, 1 point was awarded for each separate and unique example included.

Essays were scored for the number of basic essay elements by 11 professional scorers who were not familiar with the purpose or design of the study. First, a rangefinging activity was completed in order to identify the full range of responses for each element and for scoring leadership to discuss issues likely to arise during scoring. During rangefinding a sample of 50 responses across prompts were reviewed and discussed by a Scoring Director and Team Leader until consensus scores were reached for each element of each response. The responses discussed during rangefinding were then used to produce training materials. Under the guidance of the Scoring Director and Team Leader, scorers were introduced to element scoring via a scoring guide that outlined the scoring criteria. Next, scorers reviewed a set of 10 anchor essays. Following, scorers rated and discussed two, 10-response practice sets. Scorers were finally presented with an 8-response supplemental training set that consisted of the most challenging-to-classify responses. In total, scorer training lasted one and a half days.

During scoring, the Scoring Director and Team Leader reviewed scores to monitor the accuracy of scoring and check for scorer drift. Scorers were recalibrated as needed based on monitoring results. Fifteen percent of the essays were scored by scoring leadership for the
purpose of calculating agreement. Inter-rater reliability for claims, supporting reasons, elaborations, counterclaims, and conclusions were .99, .78, .64, .96, and .96, respectively.

**Writing knowledge.** Writing knowledge was measured using two open-ended items described by Harris et al. (2006). These items assessed students’ knowledge of planning and argumentative writing. Students responded to items as part of an online survey at pretest and posttest. If a student provided a response of fewer than 50 characters for either of these items, a message was displayed to think about the question(s) more and then write more ideas. Items are presented in Appendix F.

Responses were scored by dividing into idea units and crediting each unique idea, following the procedure developed by Graham, Schwartz, and MacArthur (1993). Planning knowledge was scored categorizing each idea unit as a production procedure, substantive process, motivation, seeking assistance, environmental structuring, ability, or other. Argumentative writing knowledge was scored categorizing each idea unit as a production procedure, element, or other. Writing knowledge responses were scored by two professional scorers unfamiliar with the purpose or design of the study. First, a rangefinding activity was conducted in which a sample of 50 responses for each of the two items were reviewed and discussed by scoring leadership until consensus scores were reached for each category of each response. These responses were combined into training materials in the form of a 10-response anchor set and two, 10-response practice sets. Scorers completed, reviewed, and discussed these training materials prior to scoring responses.

Fifteen percent of writing knowledge responses were additionally scored by scoring leadership for the purpose of calculating agreement. Planning knowledge interrater
reliabilities for production procedures, substantive processes, motivation, seeking assistance, environmental structuring, ability, and other were .99, .94, .99, .99, .98, 1.00, and .97, respectively. The production procedures, substantive processes, motivation, seeking assistance, environmental structuring, and ability categories accounted for 95% of all idea units produced by students. Argumentative writing knowledge interrater reliabilities for production procedures, elements, and other were .97, .90, and .93, respectively. The production procedures and elements categories accounted for 54% of all idea units produced by students.

**Writing self-efficacy.** Writing self-efficacy was measured at pretest and posttest using an adapted version of the Graham, Schwartz, and MacArthur (1993) scale. Two items from the original scale not relevant to the current study were omitted, and “paper” was changed to “essay.” Pajares, Hartley, and Valiante (2001) found writing self-efficacy measures in which students respond on a 0–100 scale to be psychometrically stronger than those using a traditional Likert scale. Since the two formats do not differ in difficulty or length, a 0–100 scale improves predictive utility (Pajares, 2003). Thus, students responded using a 0–100 scale rather than a Likert scale as recommended by Pajares (2003). Writing self-efficacy was calculated as the mean rating of all items on the scale. The instrument appears in Appendix G.

**Achievement goals.** Participants’ achievement goals were measured with the Pekrun, Elliott, and Maier (2006) revision of Elliot and McGregor’s (2001) Achievement Goals Questionnaire (AGQ). Each achievement goal (mastery approach, performance approach, and performance avoid) was assessed with three items on a scale from 1 (not at all true of
me) to 5 (very true of me). Achievement goal orientation scores were created by summing item scores for mastery-approach goals, mastery- avoidance goals, performance-approach goals, and performance-avoidance goals to create the four goal indexes. Internal consistency was $\alpha = .75, .83, 89$, and .71 for each of the four indexes, respectively. This instrument appears in Appendix H. Students’ achievement goal orientations are included in Table 3.1.

**Teachers’ writing instructional practices.** Teachers in the NC + TRAD and COMP conditions completed a survey regarding their writing instructional practices (Gilbert & Graham, 2010). Due to the scale of the study, it was not feasible to conduct classroom observations to verify these reports. However, Lane et al. (2010) and Olinghouse and Leaird (2009, as cited in Cutler & Graham, 2008) compared survey and observational data and provided evidence that teachers’ reports of writing instructional practices are generally consistent with their practices.

The first section of the survey asked teachers to provide descriptive information on their gender, ethnicity, degree(s) held, certification(s), current grade teaching writing, years spent teaching writing, the quality of their preparation to teach writing, their approach to writing instruction, and their students’ writing abilities. The second section requested information about teachers’ use of evidence-based writing practices (Graham & Perin 2007a, 2007b). To avoid biasing results, the survey did not indicate that the practices were evidence-based. The final section solicited information about writing adaptions (Cutler & Graham, 2008; Graham, Harris, MacArthur, & Fink-Chorzempa, 2003; Kiuhara, Graham, & Hawken, 2009) used for weaker writers. Information included in the survey reminded teachers that
adaptations occur only when an activity is done more often with weaker writers than it is with other students in the class. The survey of writing practices appears in Appendix I.

**Students’ and teachers’ perceptions of the events, value, and strengths and weaknesses associated with NC Write.** Students and teachers in the two treatment conditions completed surveys regarding their overall perceptions of NC Write at the end of the intervention (see Appendix J). Survey items were adopted from Grimes and Warschauer (2010). Students were asked questions related to whether they liked writing with NC Write, found the lessons in NC Write helpful, and felt NC Write helped improve their writing. Teachers were asked questions related to whether NC Write made writing instruction easier, saved them time, and let them focus on higher concerns of writing instead of mechanics. Teachers were also asked to report how much NC Write assisted or impeded the writing development of various groups of students, such as English Language Learners (ELLs). Both teachers and students were asked whether they felt the essay scores provided by NC Write were fair and whether students revised more using NC Write.

One-on-one, semi-structured interviews based on open-ended questions were used to further investigate students’ and teachers’ perceptions of the events, value, and strengths and weaknesses associated with the writing practice program. Students in both treatment conditions were asked to reflect on whether NC Write helped them improve their writing, whether they found the automated feedback useful, and what they would change about NC Write. Students in the NC + SRSD condition were additionally asked to share their opinions about the writing and self-regulatory strategies learned. Teachers in both treatment conditions were asked similar questions about NC Write, as well as about the social validity of the
intervention. Teachers in the NC + SRSD condition were additionally asked to share their opinions about the writing and self-regulatory strategies taught. Interviews were recorded and transcribed verbatim. Interview guides appear in Appendix K.

QSR International’s NVivo 10 Software was used to facilitate qualitative analysis. Qualitative data analysis strategies were borrowed from grounded theory (Strauss & Corbin, 1998) and applied to the transcribed interview data. Interview responses were initially grouped by treatment condition. The coding process first involved open coding to identify concepts, then comparative analysis to develop the dimensions and properties of concepts, followed by coding for context, process, and integration of categories (Corbin & Strauss, 2015). During this process memos and diagrams were produced and used as analytic tools. Two coders organized and coded data to build and test the coding scheme, revising and discussing until agreement was reached and categories emerged and were verified. Negative cases identified during analysis were reviewed to validate results. The final coding scheme appears in Appendix L.

Procedures

Student consent and identification. Participating teachers distributed IRB-approved informed consent forms to students, for review/approval by consenting parents. As noted above, lack of consent did not preclude students from using the program and participating in the intervention, but only data from consenting students were analyzed. Student data were stored coded with an ID number.

Pretest measures. A week prior to the start of treatment students responded to an online survey in which they rated their writing self-efficacy and completed the AGQ.
Students also responded to the two writing knowledge items. All students then wrote pretest essays in response to an assigned writing prompt as a baseline measure of writing performance.

**General instructional procedures.** Teachers in both treatment conditions received a demonstration of NC Write and were given the opportunity to try out the software prior to the start of treatment. All teachers were provided with a folder that included informed consent forms in addition to detailed directions outlining the study expectations and treatment procedures. All teachers received lessons for each session of the intervention that indicated which components of NC Write (i.e., lessons, writing/revising opportunities) should be incorporated into classroom instruction for each session. Teachers also received a document that summarized the NC Write lessons, NC Write activities, and classroom activities by session.

Teachers in both treatment conditions were additionally provided with a script to use to provide some 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 trick PEG their score would not accurately reflect the quality of their writing.

Teachers in the NC + SRSD condition were provided with background information about the SRSD model. Teachers learned about the purpose and benefits of SRSD, and were provided an overview of the six stages of instruction. The SRSD lesson plans were reviewed
with teachers, who were informed that the lesson plans should function as “metascripts” in that they should be adjusted as needed to meet students’ instructional needs. Teachers received a folder for each student containing all materials needed during the intervention, including a chart and checklist for the general planning and argumentative-specific strategies, brainstorming sheet, self-statement sheet, transition word sheet, goal sheet, and five sample essays. Each teacher was also provided with a class set of cue cards. All study materials were reviewed and discussed with teachers. All SRSD materials appear in Appendix M.

**NC + SRSD instructional procedures.** Teachers in the NC + SRSD condition implemented six recursive stages of SRSD instruction to teach students argumentative writing and self-regulatory strategies. SRSD instruction incorporated interactive NC Write lessons that emphasized self-regulatory strategies, as well as the writing practice opportunities and feedback afforded by NC Write.

**STOP and DARE.** Based on the goal of teaching middle school students strategies for argumentative writing, the STOP and DARE strategies (Graham & Harris, 2005; Harris, Graham, Mason, & Friedlander, 2008) were identified as optimal. Some adjustments were made to the DARE strategy to align with the expectations and language expressed in the CCSS for writing. Students were taught a planning strategy for argumentative writing based on the mnemonic STOP, representing Suspend judgment, Take a side, Organize ideas, and Plan more as you write. One intent of this strategy was to encourage students to choose the side of an argument they could best defend considering claims and potential counterclaims. In addition, this strategy encouraged considering how to organize this information to make a compelling argument, and continuing to plan, adjust, and revise during the writing process.
Students were additionally taught a strategy for producing the basic elements or parts of argumentative essays based on the mnemonic DARE, representing **Develop** your claim, **Add** supporting reasons and evidence, **Reject** counterclaims, **End** with a conclusion. The purpose of this strategy was to help students recall and include all of the basic argumentative essay parts in their compositions.

**Stage 1: Develop and activate background knowledge.** The first stage of instruction initiated the development and activation of background knowledge that students would need to ultimately apply the two strategies successfully. First, students were told that they would be learning how to write convincing argumentative essays. Students were asked to share examples of where they had seen argumentative essays outside of class and of times they used an argument to try to convince someone to agree with their opinion on an issue. Next, students were provided with the STOP and DARE chart and introduced to the STOP strategy. Each planning step and all terms were discussed. Students brainstormed the important parts of argumentative essays, then teachers introduced the DARE strategy and discussed each essay part. Teachers emphasized that a good argument makes sense to the reader and has several parts. Students examined model essays and identified DARE elements in each. Students rehearsed STOP and DARE and began to commit the strategies to memory. Students continued to practice recall and application of these strategies throughout the remaining stages of instruction.

**Stage 2: Discuss it.** During this stage of instruction teachers and students continued to discuss argumentative essay parts and their importance. Teachers emphasized that using STOP and DARE would help students to write good argumentative essays that are
convincing to readers. Students were introduced to the concept of establishing writing goals and were asked to consider the goal to write an argumentative essay that includes all essay parts and make sense to the reader.

Students continued to test their recall and application of STOP and DARE during this stage using a lesson in NC Write. Over the course of the lesson students completed an activity in which they were quizzed on the STOP mnemonic. Students then practiced an application of the STOP strategy via an activity that required them to classify statements that supported a claim, a related counterclaim, or both. Students were introduced to the Brainstorming to Choose a Claim graphic organizer. Finally, following a review of DARE that included examples of all essay parts, students completed an activity in which they identified essay parts given essay excerpts. For all activities, NC Write provided immediate feedback for both correct and incorrect responses. All activities were mastery-oriented in that if students provided an incorrect response they were able to change their response based on the feedback provided by NC Write until they responded successfully.

Additionally during this stage students reviewed and evaluated their pretest essays to determine which DARE essay parts they had included. Students were told not be discouraged if they included few parts since they just learned the DARE strategy. Students used the STOP and DARE checklist to record whether they included each part and the number of sentences written for each. Teachers then led a discussion of which essay parts were most frequently omitted from compositions. Students were also encouraged to consider how to improve upon the essay parts they had included, such making a claim more specific or adding reasons and evidence as support. Students were reminded how STOP and DARE would help them write
effective argumentative essays. By emphasizing the connections among effort, strategy use, and essay quality, teachers began to support students’ motivation for writing and encourage the development of adaptive attributions.

**Stage 3: Model it.** In the third stage of instruction teachers modeled all aspects of the argumentative essay writing process including both writing and self-regulatory strategies. Given the importance of collaborative cognitive modeling for effective SRSD implementation (Harris et al., 2015), teachers were provided with detailed scripts for this stage of instruction and asked to follow them closely. Teachers first reviewed the graphic organizer with students, emphasizing the role of “Suspend Judgment” in the STOP planning strategy. Students were also introduced to the STOP and DARE cue cards which they referred to throughout the remainder of modeling.

To model essay writing, teachers followed the steps for planning and writing an argumentative essay shown on the cue cards, while talking aloud throughout the process to make their thinking explicit for students. Teachers first modeled use of the STOP strategy, using the graphic organizer to generate claims and counterclaims, pick a side, and select and organize ideas to include in the essay. Students were asked to provide suggestions throughout this process. Next, teachers modeled use of the DARE strategy to include all essay parts when composing. While drafting the essay, teachers demonstrated how to continue planning while writing and made revisions “on the fly” to emphasize writing as a recursive process. They additionally modeled a wide range of self-statements, including those to related to focusing attention and defining the task (“OK, What do I need to do to get started?”), implementing strategies (“Now it is time to pick one side for my argument.”), coping (“OK, I
know I can think of good ideas if I take my time.”), self-reinforcing (“That was a lot of work, but I have a lot to show for it.”), and self-evaluating (Did I include all of the essay parts?”). Once the draft was complete, teachers acknowledged the effort and perseverance required to reiterate that an effective argumentative essay is the result of strategy use and hard work rather than innate skill. Following, teachers modeled use of the checklist to self-evaluate the writing process and essay.

To further support students’ development and use of self-statements, students completed a lesson in NC Write that provided further examples and applications of self-statements. As part of this lesson students viewed a conversation between two students that modeled use of self-statements for planning and writing an argumentative essay. Students were presented with additional examples of self-statements to use to get started as well as the reasons for using them. Students were reminded that they should be thinking of self-statements that were personalized and therefore would be useful and meaningful to them. Next, students were presented with an essay drafted by the fictional student “Carlos,” and were able to interact with various sections of the essay to see the self-statements that “Carlos” used for each while planning and writing. Students then viewed a second conversation between two students that modeled coping and self-reinforcement self-statements, and were provided with additional examples of these types of self-statements. Finally, students interacted with these self-statements to learn more about the importance of noticing effective strategy use. Following the lesson, students recorded their own self-statements to use while planning, composing, and evaluating essays on their self-statements sheet. Students continued to add self-statements to this sheet over the course of the remaining stages.
**Stage 4: Memorize it.** Students practiced recalling the STOP and DARE mnemonics, as well as all steps for planning and writing an argumentative essay, throughout the intervention. A few minutes of each writing session was allocated for students to practice memorization and teachers offered a range of activities to help students process the strategies and commit them memory. Individual students were provided with additional or differentiated opportunities to practice as needed. Teachers regularly assessed students’ knowledge of the mnemonics and the strategy steps.

**Stage 5: Support it.** During the fifth stage of instruction the primary responsibility for essay writing shifted from teacher to student. However, students’ use of the writing and self-regulatory strategies continued to be supported by their teacher and peers. Additional scaffolds employed throughout this stage included the students’ charts, checklists, cue cards, graphic organizer, and self-statements sheet. These scaffolds were faded as appropriate.

This stage began with a collaborative essay writing activity. Teachers and students set a goal to include all four essay parts, as well as five “ideas” (three arguments for the claim and two disputed counterclaims) in the essay. Students recorded these goals on their goal sheets. Referring to their charts, cue cards, and graphic organizer, students led the process of planning and writing a collaborative essay, with teachers providing assistance as needed. Students populated their own organizers, and then teachers facilitated a discussion to determine which side to pick based on the available evidence. The class reached consensus about the best ideas to include and a logical ordering. Finally, students volunteered one sentence at a time to build essays. Teachers recorded essays on the board or chart paper while students copied on notebook paper.
The following session students were referred to their goal sheets and reminded of their goals for the essay, and teachers made any adjustments to goals for individual students. Students were challenged to find at least two to three ways to improve the essays. Students then transcribed essays into NC Write, making edits during the process. Students were reminded to refer to their self-statements while writing. When students had completed their essays they used their checklists to evaluate for all essay parts prior to submitting. Following, students reviewed the scores and feedback provided by PEG. Teachers addressed questions with individual students, helping them to understand how their essays could be further improved using the feedback provided by PEG. At this time some students may have believed that including all essay parts was sufficient to earn a top score; teachers seized this opportunity to help students understand that it was not just the presence of essay parts but also the quality which determined the strength of their argument and the overall quality of their essay. Students compared their completed essays to their pretest essays, and shared ways that their recent essays were better.

As an additional support during this stage, students worked in small groups to revise a sample essay of mediocre quality. Each group’s goal was to ensure that the revised essay contained all essay parts, and that all parts were of a high quality. Students referred to their cue cards during the writing process. Groups shared their completed essays at the conclusion of this activity.

Throughout the remainder of this stage students composed, on average, four more essays; set goals for each essay that included a total score along with the number of essay parts; evaluated each essay for number of parts; and revised each essay considering how the
PEG feedback related to the presence and quality of essay parts. Teachers emphasized the importance of using the feedback to make substantive revisions that improve the development, organization, and/or structure of the essay, rather than only surface-level revisions such as spelling and conventions. Students also completed four additional NC Write lessons during this time. As students developed proficiency with the strategies teachers faded support. To encourage generalization of strategies teachers periodically asked students to share instances they had thought about or used STOP and DARE outside of the classroom.

**Stage 6: Independent performance.**

The sixth and final stage required students to plan and write an argumentative essay independently, without referencing the resources used during previous stages or receiving feedback from the teacher or PEG. Posttest essays functioned as measures of students’ independent writing performance.

**NC Write + TRAD instructional procedures.** Teachers in the NC Write + TRAD condition implemented their own writing instruction. This traditional writing instruction incorporated interactive NC Write lessons focused on the development of ideas, organization, word choice, sentence structure, and conventions. Students in this condition had access to the same writing practice opportunities and feedback afforded by NC Write as students in the NC + SRSD condition. Students in both conditions completed an equal number of NC Write lessons though the content of these lessons differed.

Results of the survey of writing practices indicated that half of the teachers in this condition (three out of six) received no to minimal writing preparation during their teacher certification programs, while the other half of teachers received adequate preparation. Five
teachers, however, received adequate to extensive writing preparation after certification while one teacher received no preparation. Finally, half of the teachers rated their personal writing preparation as minimal and the other half as adequate.

Results of the survey of writing practices further indicated that teachers in the NC + TRAD condition used a combination of process and skills approaches to writing instruction. Direct instruction of skills and summarization instruction were the most frequently reported evidence-based practices, which all teachers indicated using on at least a weekly basis. Other high-frequency practices reported by all teachers included using writing as a learning tool, which five out of six teachers indicated using at least weekly, and paragraph writing, which half the teachers reported using weekly or more often. All teachers also reported teaching spelling and sentence combining. Half of the teachers indicated these practices were used on a weekly or more frequent basis. All teachers reported teaching inquiry/research, with four teachers indicating this occurred at least monthly. All teachers indicated that students frequently used word processing. Only two teachers reported teaching any handwriting or typing.

Considering aspects of instruction emphasized in the NC + SRSD condition, all teachers reported teaching planning strategies and using student self-assessment; only half of the teachers reported using these practices on a weekly or more frequent basis. All teachers indicated teaching strategies for revising, though only two teachers did so weekly. All teachers also indicated teaching students to imitate models of good writing and half of the teachers used this practice on a monthly or more frequent basis. All teachers reported using prewriting activities and peer collaboration to plan, draft, or revise, and most teachers
indicated these were used frequently. All teachers reported using teacher-set writing goals and four teachers indicated such goals were used regularly.

The most frequently provided adaptation for weaker writers across the NC + TRAD classrooms was extra time to do assignments, which all teachers reported providing weekly or more often. All teachers also reported providing extra time to practice skills/strategies, which five out of six teachers indicated using at least weekly. Other frequently used adaptations included extra review of skills/strategies, conferencing, and opportunities to use word processing, which were reported by all teachers and used weekly or more often by four of the teachers. All teachers also reported providing extra sentence writing instruction, grammar instruction, mini-lessons, and opportunities to write with peers, of which half the teachers indicated using on at least a weekly basis. Less frequently used adaptations included extra revising instruction, planning instruction, re-teaching of skills/strategies, and permitting students to select their own writing topic, which all teachers used but only two teachers reported using weekly or more often. All teachers reported providing alternative writing assignments to students on a monthly or more frequent basis, and all reported providing extra spelling instruction, with four teachers indicating this was used at least monthly. Finally, five teachers reported providing extra writing instruction via technology as an adaption for weaker writers and three of these teachers indicated this occurred weekly or more often.

Teachers in the NC + TRAD condition elaborated on their writing practices during interviews, in which they were asked to describe their approach to teaching argumentative writing. Five of the six teachers indicated that writing has become less of a priority following the elimination of the NC Writing Assessment Program formerly administered in grades 4, 7,
and 10. As a result, teachers described increasing the integration of reading and writing, incorporating elements of argumentative writing into research papers, and utilizing abbreviated writing sessions. A number of teachers felt they took a “formulaic” approach to argumentative writing instruction, teaching a five paragraph essay structure. In general writing skills were primary taught through mini-lessons several times a week. Regarding aspects of instruction emphasized in the NC + SRSD condition, while some teachers described modeling use of graphic organizers to plan, only one teacher described teaching students to consider the quantity/quality of reasons and evidence as part of planning. Teachers varied in their approach to teaching argumentative essay elements. Most emphasized claim, reasons and evidence, and conclusion, and only one teacher described using sample essays to provide examples of essay elements. Finally, no teachers described teaching self-regulatory skills such as goal-setting, self-instructions, or self-reinforcement.

Teachers in the NC + TRAD condition implemented their own writing instruction to teach students the basic structure of argumentative essays and a variety of relevant writing skills. Though writing instruction in this condition overlapped with writing instruction in the NC + SRSD condition in terms of content (e.g., both taught how to use transition words to improve essay organization), survey and interview results indicated that traditional writing instruction did not emphasize strategies for planning, writing, and revising argumentative essays to the extent these were emphasized in the NC + SRSD condition. Additionally, teachers in NC + TRAD condition did not revisit strategies regularly until students demonstrated mastery and teach students self-regulatory skills as was the case in the NC + SRSD condition.
COMP instructional procedures. Students in the COMP condition received traditional writing instruction from their regular teacher. Results of the survey of writing practices indicated that most teachers in this condition (four out of five) received minimal to adequate writing preparation during their teacher certification programs and after certification. However, one teacher reported extensive writing preparation both during and after certification. Additionally, four out of five teachers rated their personal writing preparation as adequate, while one teacher rated it as extensive.

Previous conversations with teachers and school administrators indicated that COMP condition teachers provided a comparable amount of writing instruction to teachers in the treatment conditions (i.e., at least 60–90 minutes per week). Results of the survey of writing practices indicated that COMP condition teachers used a combination of process and skills approaches to writing instruction. Summarization instruction was the most frequently reported evidence-based practice which all teachers used and four out of five teachers reported using at least monthly. All teachers also reported providing direct instruction of writing skills, paragraph writing instruction, and teaching students to imitate models of good writing. Three out of five teachers indicated these practices were used on a monthly or more frequent basis. Additionally, all teachers reported using writing as a learning tool, with two out of five indicating this was used on at least a weekly basis. Four out of five teachers reported teaching sentence combining, inquiry/research, and having students use word processing, though on average, these practices were not used as frequently as those mentioned above. Finally, a single teacher reported teaching handwriting and none typing.
Considering aspects of instruction emphasized in the NC + SRSD condition, all teachers reported teaching students revising strategy instruction and to imitate models of good writing; three out of five teachers reported using these practices on a monthly or more frequent basis. All teachers also reported teaching planning strategies and student-self assessment, though only two teachers reported using the former and one the latter on a monthly or more frequent basis. Further, though all teachers reported using teacher-set writing goals only one teacher indicated using more than occasionally. Four out of five teachers reported using peer collaboration to plan, draft, or revise, though none of the four more than occasionally. Finally, four out of five teachers indicated using prewriting activities though only one teacher did so more than occasionally.

Regarding adaptations used with weaker writers, all teachers in the COMP condition reported providing extra time to complete assignments, extra sentence writing instruction, and extra conferencing, with three out of five teachers indicating these adaptations were used on at least a monthly basis. All teachers also reported providing extra review of skills/strategies; two teachers indicated this occurred on at least a weekly basis. Four out of five teachers reported providing extra time to practice skills/strategies, and three of the teachers indicated doing so on at least a monthly basis. Additionally, four out of five teachers reported providing extra grammar instruction, opportunities to write with peers, planning instruction, and opportunities to use word processing, in addition to re-teaching skills/strategies, providing alternative writing assignments, and allowing students to select their own topic. However, only two of the teachers reported using these adaptations on at least a monthly basis. Finally, only three out of five teachers reported using extra mini-
lessons and writing instruction via technology as adaptations, and two of the teachers
indicated using these strategies at least monthly. Only three out of five teachers also reported
providing extra revising instruction as an adaptation, though only one of the teachers
indicated using this on a monthly basis.

Despite some differences in the frequency of the various writing practices taught and
adaptations used between teachers in the COMP condition, results of the writing practice
survey indicated that in general COMP condition teachers took an approach to instruction
that substantially differed from the SRSD approach. Specifically, these teachers did not
emphasize strategies for planning, writing, and revising argumentative essays, revisit
strategies regularly until students demonstrated mastery, and focus on self-regulatory skills as
was the case in the NC + SRSD condition.

**Posttest measures.** The week following the conclusion of treatment students once
again responded to an online survey in which they rated their writing self-efficacy and
responded to the two writing knowledge items. All students then wrote posttest essays in
response to an assigned writing prompt as a post-treatment measure of writing performance.

**Fidelity of treatment implementation.** Several steps were taken to ensure and
measure the fidelity of the treatments implemented in the NC + SRSD and the NC + TRAD
conditions. First, teachers in both conditions received detailed directions outlining the study
expectations and treatment procedures. Teachers also received lessons for each day of the
intervention. Teachers were provided with all materials that they and their students required
for the intervention. Second, teachers received a demonstration of NC Write, and were given
the opportunity to try out the software prior to the start of treatment. Third, teachers were
invited to contact the author directly with any questions. Additionally, each teacher was contacted weekly regarding progress and any recent issues encountered. Reported issues were uncommon and tended to involve either NC Write accounts (e.g., establishing accounts for new students) or helping students in the NC + SRSD condition accurately self-evaluate their essays. Solutions to these issues were discussed and implemented. Fourth, NC Write provided a measure of standardization for the lessons and essay writing and documented implementation fidelity. For example, each student who completed a given lesson in NC Write was guaranteed to be presented with the same content. All essays written and revised were logged in NC Write. As checklists of lesson steps or activities provide only self-reported information about teachers and no information regarding individual student activity teachers were not asked to complete such measures. Instead, data collected by NC Write were mined to evaluate implementation fidelity. Using program logs overcame limitations of checklists commonly used in intervention research which report only class-level data. It is well known that self-reported data are subject to social desirability and other biases. Further, teacher reports of intervention steps completed provide no information about whether a given student completed each activity or was even present the day of the activity. Program logs provided data about each student’s lessons accessed and essays written in NC Write. Lesson activity logs indicated both whether students had accessed each intended lesson and how long they spent completing each lesson. Essay writing logs showed each essay written and revised by date and prompt. Analysis of the lesson activity and essay writing logs revealed that students in the NC + SRSD condition completed 81% of treatment activities, on average,
and that students in the NC + TRAD condition completed 76% of treatment activities, on average.

Finally, an intervention fidelity variable was created ($M = 13.31, SD = 2.64$) that was a count of the NC Write lessons, initial drafts, and revisions each student completed over the course of the intervention. Both lesson log data in NC Write and all essays were reviewed to determine the validity of lesson completion and originality of essays. Any lessons that students accessed for fewer than two minutes in total and essays submitted with plagiarized text were excluded from these counts. Independent samples $t$-test results revealed significant differences in intervention fidelity scores between the NC + SRSD ($M = 13.73, SD = 2.73$) and NC + TRAD ($M = 12.86, SD = 2.48$) conditions; $t(557) = 3.93, p < .01$.

**Interviews.** In the second phase of the study, interviews were conducted one-on-one with students and teachers with the aim of further understanding their perceptions of the events, value, and strengths and weaknesses associated with NC Write.

**Subjectivity Statement**

My entrance to the education profession was through a lateral entry program following a few years of substitute teaching. Most of my teaching experience has been in the North Carolina public school system teaching upper-elementary grades. The challenge of how to best support students’ writing development has intrigued me from the time I started teaching, particularly as I considered the complexity of writing and experienced firsthand the competing demands on teachers’ instructional time. The NC Writing Assessment Program was still in place when I was teaching and it seemed writing received greater instructional emphasis than it does currently (though even then it seemed there was never enough time to
provide students with sufficient practice and feedback). During my graduate work I had the opportunity to teach both undergraduate and graduate courses at North Carolina State University, which provided me with additional, albeit anecdotal, evidence of the generally inadequate quality of students’ writing. Experiences working with students and exposure to research during the course of my doctoral studies have produced in me strong beliefs about the value of self-regulation and concerns related to an over-emphasis on the individual differences of students.

I eventually left the classroom and pursued a career in the educational assessment industry. Through experiences associated with this career (more so than teaching) I came to better understand the importance of using high-quality assessments, including formative assessments, to support valid inferences about student knowledge and abilities. At the time of this writing my day-to-day work involves the management and operations oversight of multi-state, large-scale assessment programs. Working with state departments of education across the country has provided insight into the challenges related to assessment validity, cost, and reporting turnaround time with which policymakers struggle. My employer, Measurement Incorporated, developed and licenses NC Write. I was not involved with the development of NC Write outside of the SRSD lessons developed specifically for the present study. Measurement Incorporated provided me with unconditional permission to use NC Write for the purpose of the present study.
CHAPTER FOUR

Results

The present study examined changes in students’ writing performance, knowledge, and self-efficacy associated with NC Write + SRSD instruction and NC Write + traditional writing instruction compared to traditional writing instruction. The shape of growth in students’ writing performance associated with these two treatments was also investigated. Students’ and teachers’ perceptions of the events, value, and strengths and weaknesses associated with the writing program were examined using both quantitative and qualitative data. This section contains both descriptive statistics and analyses for the study’s primary research questions.

Prior to conducting analyses intervention fidelity measures were examined to identify cases to exclude from analyses. Students were excluded from analysis if they met any one of the following conditions: changed classes during the course of the intervention, did not complete the pretest and/or posttest essays, or completed fewer than three essays during the intervention (not including the pretest/posttest essays). Additionally, as NC Write did not include plagiarism detection at the time of the study, all essays were screened for plagiarism and plagiarized essays were excluded from all essay counts and analyses. Survey completion times at pre- and posttest were examined to screen for outlier values (−3 SDs), i.e., students who completed the survey too quickly to answer questions meaningfully. No outlier times were identified.
Preliminary Analyses

Graham, Schwartz, and MacArthur (1993) did not examine the psychometric properties of the self-efficacy measure adapted in the present study. Therefore, exploratory factor analyses were conducted on the self-efficacy pretest items and scale separately for each timepoint. An initial examination of item inter-correlations revealed that the Pearson correlation coefficients among the regular (i.e., non-reverse-coded) items ranged from .43 to .61 on the pretest and .28 to .55 on the posttest. Similarly, correlations among the reverse-coded items ranged from .36 to .42 on the pretest and .40 to .45 on the posttest. However, the regular and reverse-coded items ranged from not correlated to weakly correlated with each other (Pearson correlation coefficients ranging from −.07 to .03 on the pretest and to −.02 to .38 on the posttest). None of the reverse-coded pretest items were significantly positively correlated with the regular items, and though all of the reverse-coded posttest items were significantly positively correlated with one of the five regular items, only one of these exceeded the .3 level. This suggested that the 8-item self-efficacy instrument did not assess the expected unified construct. The reverse-coded items were omitted from the analysis. The remaining five items were first screened for outlier values (there were none at pre- or posttest), then evaluated to determine suitability for exploratory factor analysis.

For the five-item scale the Kaiser-Meyer-Olkin measure of sampling adequacy was great for both the pretest (.85) and posttest (.82) data. Bartlett’s test of sphericity was significant for both the pretest ($\chi^2 (10) = 1503.47, p < .001$) and posttest ($\chi^2 (10) = 980.11, p < .001$). Thus, the items were determined to be appropriate for factor analysis. Principle components extraction on the data for each time point resulted in a single-factor solution that
explained 62.53% and 53.21% of the total pretest and posttest variance, respectively. All items at both time points had loadings >.5 with the exception of item 3 which had a loading of .3 at posttest. This item was retained as it was part of the original instrument and had a loading of .59 at pretest. The factor loadings for both time points are presented in Table 4.1. Internal consistency of the five-item scale was $\alpha = .85$ at pretest and $\alpha = .77$ at posttest.

<table>
<thead>
<tr>
<th>Item</th>
<th>Pretest Factor 1 Loading</th>
<th>Posttest Factor 1 Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. When writing an essay it is easy for me to get ideas.</td>
<td>.66</td>
<td>.62</td>
</tr>
<tr>
<td>3. When writing an essay it is easy for me to get started.</td>
<td>.59</td>
<td>.30</td>
</tr>
<tr>
<td>4. When writing an essay I find it easy to make all the changes I need to make.</td>
<td>.63</td>
<td>.52</td>
</tr>
<tr>
<td>5. When writing an essay it is easy for me to write my ideas into good sentences.</td>
<td>.69</td>
<td>.67</td>
</tr>
<tr>
<td>8. When my class is asked to write an essay, mine is one of the best.</td>
<td>.56</td>
<td>.54</td>
</tr>
</tbody>
</table>

*Note. N = 771 pretest, 774 posttest*

A confirmatory factor analysis was conducted with AMOS 21 (Arbuckle, 2012) to confirm that the AGQ displayed the same structural properties as the original instrument: mastery-approach goals, mastery avoid goals, performance-approach goals, and performance- avoidance goals. Of the total sample of students, 771 (93%) completed the AGQ. There were no missing data for these students. Assumptions of multivariate normality and linearity were evaluated with SPSS 23.0 using box plots and Mahalanobis distance. This resulted in the removal of 5 outlier cases with Mahalanobis distance values at or above 19.52. Covariance matrices served as input and solutions were generated using maximum-likelihood estimation. Multiple criteria including absolute and incremental fit indexes were used to
evaluate model fit. As recommended by Hu and Bentler (1999), acceptable model fit was defined as a comparative fit index (CFI) ≥ .95; Tucker–Lewis index (TLI) ≥ .95; and root mean square error of approximation (RMSEA) ≤ .06, 90% confidence interval (CI) ≤ .06, test of close fit (CFit) ns. Results confirmed that the hypothesized four-factor solution had an acceptable fit to the data, \( \chi^2(48, N = 766) = 174.82, p < .001; \) CFI = .98; TLI = .98; RMSEA = .059 (90% CI = .050–.068, CFit = .06). All factor loadings were significant, and the average factor loading was .75. The hypothesized four-factor solution fit significantly better than a three-factor model comprising performance-approach and -avoidance goals and omnibus mastery goals (mastery-approach and -avoidance goals collapsed together), \( \Delta \chi^2(3) = 456.42, p < .001, \) and better than a three-factor model comprising performance-approach and mastery-approach goals and omnibus avoidance goals (performance- and mastery-avoidance goals collapsed together), \( \Delta \chi^2(3) = 661.64, p < .001. \) Results from all models are displayed in Table 4.2.

Table 4.2. Results from AGQ Confirmatory Factor Analyses

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \chi^2(N = 766) )</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesized Model</td>
<td>( \chi^2(48) = 174.82 )</td>
<td>.98</td>
<td>.98</td>
<td>.059</td>
</tr>
<tr>
<td>Omnibus Mastery Model</td>
<td>( \chi^2(51) = 631.24 )</td>
<td>.92</td>
<td>.90</td>
<td>.122</td>
</tr>
<tr>
<td>Omnibus Avoidance Model</td>
<td>( \chi^2(51) = 836.46 )</td>
<td>.89</td>
<td>.86</td>
<td>.142</td>
</tr>
</tbody>
</table>

*Note. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation*

**Main Analyses Overview**

Descriptive statistics for pre- and posttest assessments of writing quality, knowledge, and self-efficacy by condition are displayed in Table 4.3.
Table 4.3. Means and standard deviations for study variables by condition and time of testing

<table>
<thead>
<tr>
<th>Variable</th>
<th>Condition</th>
<th>NC Write + traditional writing instruction</th>
<th>NC Write + SRSD instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
<td>Pretest</td>
</tr>
<tr>
<td>Writing performance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td>10.67</td>
<td>9.93</td>
<td>13.03</td>
</tr>
<tr>
<td></td>
<td>(3.72)</td>
<td>(3.87)</td>
<td>(3.81)</td>
</tr>
<tr>
<td>Essay length</td>
<td>166.42</td>
<td>140.97</td>
<td>245.49</td>
</tr>
<tr>
<td></td>
<td>(119.03)</td>
<td>(110.67)</td>
<td>(154.87)</td>
</tr>
<tr>
<td>Essay elements</td>
<td>6.48</td>
<td>5.93</td>
<td>8.44</td>
</tr>
<tr>
<td></td>
<td>(3.47)</td>
<td>(3.42)</td>
<td>(4.32)</td>
</tr>
<tr>
<td>Writing knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Planning knowledge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prod. proced.</td>
<td>.04</td>
<td>.05</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>(.24)</td>
<td>(.26)</td>
<td>(.41)</td>
</tr>
<tr>
<td>Sub. process.</td>
<td>1.86</td>
<td>2.16</td>
<td>1.91</td>
</tr>
<tr>
<td></td>
<td>(1.02)</td>
<td>(1.34)</td>
<td>(1.09)</td>
</tr>
<tr>
<td>Motivation</td>
<td>.02</td>
<td>.00</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>(.13)</td>
<td>(.07)</td>
<td>(.12)</td>
</tr>
<tr>
<td>Seeking assistance</td>
<td>.25</td>
<td>.18</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>(.59)</td>
<td>(.62)</td>
<td>(.58)</td>
</tr>
<tr>
<td>Environ. struct.</td>
<td>.06</td>
<td>.09</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>(.27)</td>
<td>(.53)</td>
<td>(.33)</td>
</tr>
<tr>
<td>Ability</td>
<td>.02</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>(.17)</td>
<td>(.07)</td>
<td>(.12)</td>
</tr>
<tr>
<td>Arg. writing knowl.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prod. proced.</td>
<td>.28</td>
<td>.22</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>(.80)</td>
<td>(.74)</td>
<td>(.62)</td>
</tr>
<tr>
<td>Elements</td>
<td>.83</td>
<td>.99</td>
<td>1.27</td>
</tr>
<tr>
<td></td>
<td>(1.08)</td>
<td>(1.17)</td>
<td>(1.18)</td>
</tr>
<tr>
<td>Writing self-efficacy</td>
<td>53.63</td>
<td>52.11</td>
<td>54.92</td>
</tr>
<tr>
<td></td>
<td>(17.92)</td>
<td>(22.07)</td>
<td>(16.95)</td>
</tr>
</tbody>
</table>

Note. N = 829. The writing knowledge and self-efficacy data are from participants who completed the measures both at pretest and posttest.


*Argumentative writing knowledge.

Multilevel modeling was used to answer research questions one and two. Such a framework can account for non-independence of observations (McCoach, 2010), such as when there are multiple occasions per student or when students are clustered with teachers as they were in the present study. Further, balanced data are not required; in unbalanced datasets,
multilevel modeling makes use of all available data for each student and controls for the
timing of the measurement (Karney & Bradbury, 1997). Finally, multilevel modeling allows
for partitioning sources of variability in order to explain relations that vary by cluster, such as
within and between teachers. All models were built using SAS software version 9.3 with
Maximum Likelihood estimation. All estimated models met convergence criteria.

**RQ1: What changes in students’ writing performance, knowledge, and self-efficacy are
associated with NC Write + SRSD instruction and NC Write + traditional writing
instruction compared to traditional writing instruction?**

The following model was used to examine changes in students’ writing performance,
knowledge, and self-efficacy associated with NC Write + SRSD instruction and NC Write +
traditional writing instruction compared to traditional writing instruction:

\[
Y_{ij} = \beta_{0ij} + \beta_{1ij}(\text{Pretest}_{ij} - \overline{\text{Pretest}_{ij}}) + \beta_{2ij}(\text{NC+Traditional}) + \beta_{3ij}(\text{NC+SRSD}) + r_{ij}
\]

\[
\beta_{0j} = \gamma_{00} + u_{0j}
\]

\[
\beta_{1j} = \gamma_{10}
\]

\[
\beta_{2j} = \gamma_{20}
\]

\[
\beta_{3j} = \gamma_{30}.
\]

At Level 1, \(Y_{ij}\) represents the posttest outcome for (1) writing performance (i.e.,
writing quality, essay length, essay elements), (2) writing knowledge (i.e., planning
knowledge, argumentative writing knowledge), and (3) writing self-efficacy, predicted for
student \(i\) with teacher \(j\), \(\beta_{0ij}\) represents the intercept of the predicted posttest outcome, \(\beta_{1ij}\)
represents a control for pretest performance on the posttest outcome, \(\beta_{2ij}\) and \(\beta_{3ij}\) represent
condition assignment (1 = in condition, 0 = not in condition), and $r_{ij}$ represents the variance in the predicted posttest outcome for student $i$ with teacher $j$.

At Level 2, the individual intercepts become the outcomes ($\beta_{0j}$). Additionally at Level 2, $\gamma_{00}$ represents the predicted outcome for a student in the COMP condition holding constant pretest performance and $u_{0j}$ represents the variance around the mean of the posttest outcome, $\gamma_{10}$ represents the pretest outcome slope, or predicted change in posttest outcome based on unit change in pretest performance for the sample, $\gamma_{20}$ represents the NC + TRAD condition slope, or difference in posttest outcome predicted for students in the NC + TRAD condition controlling for pretest performance, and $\gamma_{30}$ represents the NC + SRSD condition slope. To make the intercept more interpretable all continuous pretest variables (writing quality, essay length, essay elements, planning knowledge, argumentative writing knowledge, writing self-efficacy) were grand mean centered.

Initial analyses were conducted for each outcome variable using a fully unconditional model (Model 1) that did not include any predictors. The purpose of these analyses was to ensure there was sufficient variability between and within teachers to support the use of a two-level model. The results of all models indicated that there was sufficient variability at both levels to warrant further analysis (see Tables 4.4–4.9).

**Writing performance.**

**Writing quality.** Following the unconditional model for writing quality, a second model (Model 2) was estimated to control for pretest writing quality and include treatment variables as predictors. Allowing students’ pretest writing quality slopes to vary randomly provided better model fit than constraining the slopes, ($\chi^2 (2) = 6.5, p < .05$). Results showed
that, averaging across students and within teachers, students produced higher-quality essays at posttest than at pretest ($\gamma_{10} = 0.50, t = 12.67, p < .001$). Additionally, students in the NC Write + TRAD produced higher-quality posttest essays than students in the COMP condition ($\gamma_{20} = 2.41, t = 2.12, p < .05$). Further, students in the NC + SRSD condition produced the highest-quality essays at posttest ($\gamma_{30} = 3.95, t = 3.56, p < .01$). This model explained 23% of the within-teacher variability in posttest writing quality.

Finally, a third model (Model 3) was estimated to further explain the random variation in posttest writing quality. This model added controls for gender (female = 1), age, race/ethnicity (groups included Black; Hispanic; and a composite group of Asian, American Indian or Alaska Native, Native Hawaiian or other Pacific Islander, and two or more races), Limited English Proficiency (LEP) status, and Students with Disabilities (SWD) status. A control was also included for previous ELA achievement, based on students’ scale scores (range: 423–484) from the previous year’s end-of-grade assessment. Continuous variables were grand mean centered so that between-teacher differences in students’ mean values contributed to parameter estimates. Successive models were evaluated using these variables as predictors of the intercept. Age and race/ethnicity were found to be nonsignificant predictors of posttest writing quality and were removed from the final model for parsimony. Results of the final model showed that, averaging across students, within teachers, and controlling for student demographics and pretest writing quality, students in the NC + TRAD condition produced higher-quality posttest essays than students in the COMP condition by approximately 2.9 points ($\gamma_{20} = 2.87, t = 2.33, p < .05$), while students in the NC + SRSD condition produced higher-quality posttest essays than those produced by students in the
COMP condition by over four points ($\gamma_{30} = 4.15, t = 3.49, p < .01$). This model explained 33% of the within-teacher variability in posttest writing quality. Results of all models are presented in Table 4.4.

Table 4.4.  
*Unstandardized Coefficients (and Standard Errors) of Multilevel Models of Writing Quality*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Unconditional model</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest writing quality, $\beta_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>13.27 (.69)***</td>
<td>10.88 (.86)***</td>
<td>10.47 (.97)***</td>
</tr>
<tr>
<td>Pretest writing quality, $\beta_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{10}$</td>
<td>.50 (.04)***</td>
<td>.32 (.05)***</td>
<td></td>
</tr>
<tr>
<td>NC + TRAD condition, $\beta_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{20}$</td>
<td>2.41 (1.14)*</td>
<td>2.87 (1.24)*</td>
<td></td>
</tr>
<tr>
<td>NC + SRSD condition, $\beta_3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{30}$</td>
<td>3.95 (1.11)**</td>
<td>4.15 (1.19)**</td>
<td></td>
</tr>
<tr>
<td>Gender, $\beta_4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{40}$</td>
<td>.77 (.22)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEP, $\beta_5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{50}$</td>
<td>1.91 (.54)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWD, $\beta_6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{60}$</td>
<td>−.82 (.41)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous ELA achievement, $\beta_7$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{70}$</td>
<td>.10 (.01)***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest writing quality ($\tau_{00}$)</td>
<td>8.61 (2.88)***</td>
<td>3.68 (1.31)***</td>
<td>4.01 (1.46)***</td>
</tr>
<tr>
<td>Pretest writing quality slope ($\tau_{11}$)</td>
<td>.02 (.01)</td>
<td>.02 (.01)</td>
<td></td>
</tr>
<tr>
<td>Within-teacher variance ($\sigma^2$)</td>
<td>9.61 (.48)***</td>
<td>7.41 (.37)***</td>
<td>6.46 (.38)***</td>
</tr>
<tr>
<td>Deviance (-2LL)</td>
<td>4295.7</td>
<td>4086.3</td>
<td>2961.2</td>
</tr>
<tr>
<td>Number of est. parameters</td>
<td>3</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>

*Note. *$p<.05$, **$p<.01$, ***$p<.001$*
**Essay length.** Following the unconditional model for essay length, a second model (Model 2) was estimated to control for pretest essay length and include treatment variables as predictors. Allowing the NC + TRAD condition slopes to vary randomly provided better model fit than constraining the slopes, ($\chi^2 (1) = 5.2, p < .05$). Results showed that, averaging across students and within teachers, students wrote longer essays at posttest than at pretest ($\gamma_{10} = 0.48, t = 13.10, p < .001$). Controlling for pretest essay length, students in the NC + TRAD condition wrote non-significantly longer essays at posttest than students in the COMP condition ($\gamma_{20} = 96.14, t = 1.83, p = .09$). Students in the NC + SRSD condition wrote the longest essays at posttest ($\gamma_{30} = 121.18, t = 4.07, p < .001$). This model explained 16% of the within-teacher variability in posttest length.

Finally, successive third models (Model 3) were estimated to further explain the random variation in posttest essay length by adding controls for demographic variables and previous ELA achievement. Age, race/ethnicity, and SWD status were nonsignificant predictors of posttest essay length and were removed from the final model for parsimony. Results of the final model showed that, averaging across students, within teachers, and controlling for student demographics and pretest essay length, students in the NC + TRAD condition produced essays that were approximately 99 words longer than those produced by students in the COMP condition at posttest, a non-significant difference ($\gamma_{20} = 98.95, t = 1.72, p = .10$). Students in the NC + SRSD condition produced essays that were approximately 121 words longer than those produced by students in the COMP condition at posttest ($\gamma_{30} = 120.79, t = 4.33, p < .001$). This model explained 9% of the within-teacher variability in posttest essay length. Results of all models are presented in Table 4.5.
### Table 4.5.

| Unstandardized Coefficients (and Standard Errors) of Multilevel Models of Essay Length |
|---------------------------------|------------------|------------------|------------------|
| **Fixed Effects**               | **Model 2**      | **Model 3**      |
| **Unconditional model**         | **Model 2**      | **Model 3**      |
| Posttest essay length, $\beta_0$ | 247.53 (24.72)*** | 167.35 (22.87)*** | 155.45 (22.63)*** |
| Intercept, $\gamma_{00}$        |                  |                  |
| Pretest essay length, $\beta_1$ | .48 (.04)***     | .37 (.05)***     |
| Intercept, $\gamma_{10}$        | 96.14 (52.45)    | 98.95 (57.56)    |
| NC + TRAD condition, $\beta_2$  |                  |                  |
| Intercept, $\gamma_{20}$        | 121.18 (29.76)***| 120.79 (27.91)***|
| NC + SRSD condition, $\beta_3$  |                  |                  |
| Intercept, $\gamma_{30}$        |                  |                  |
| Gender, $\beta_4$               |                  |                  |
| Intercept, $\gamma_{40}$        | 29.29 (9.04)**   |
| LEP, $\beta_5$                  |                  |                  |
| Intercept, $\gamma_{50}$        | 53.08 (21.53)*   |
| Previous ELA achievement, $\beta_6$ |                  |                  |
| Intercept, $\gamma_{60}$        | 2.96 (.46)***    |
| **Random Effects**              |                  |                  |
| Posttest essay length ($\tau_{00}$) | 11176.00         | 2399.80          | 1823.72          |
| (3745.44)***                    | (1213.67)*       | (1005.91)*       |
| NC + TRAD condition slope ($\tau_{11}$) | 10676.00         | 14687.00         |
| (7795.15)                       | (9830.12)        |
| Within-teacher variance ($\sigma^2$) | 12353.00         | 10361.00         | 11227.00         |
| (613.74)***                     | (516.14)***      | (642.80)***      |
| Deviance (-2LL)                 | 10230.5          | 10058.4          | 7720.2           |
| Number of est. parameters       | 3                | 7                | 10               |

*Note.* *p<.05, **p<.01, ***p<.001

**Essay elements.** Following the unconditional model for essay elements, a second model (Model 2) was estimated to control for pretest elements and include treatment variables as predictors. Allowing the NC + TRAD condition slopes to vary randomly provided better model fit than constraining the slopes, ($\chi^2(1) = 4.8, p < .05$). Results showed that, averaging across students and within teachers, students included more elements at
posttest than pretest ($\gamma_{10} = 0.54, t = 14.99, p < .001$). However, there were no significant differences between the number of elements included by students in the NC + TRAD and COMP conditions ($\gamma_{20} = 2.11, t = 1.52, p = .15$). Students in the NC + SRSD condition included significantly more basic elements of argumentative essays in their posttest essays than students in either of the other conditions ($\gamma_{30} = 2.68, t = 3.31, p < .01$). This model explained 20% of the within-teacher variability in posttest essay elements.

Finally, successive third models (Model 3) were estimated to further explain the random variation in posttest essay elements by adding controls for demographic variables and previous ELA achievement. Age, race/ethnicity, LEP, and SWD status were found to be nonsignificant predictors of posttest essay elements and were removed from the final model for parsimony. Results of the final model showed that, averaging across students, within teachers, and controlling for student demographics and pretest essay elements, students in the NC + TRAD condition included approximately 2.2 more elements than students in the COMP condition in their essays at posttest, a non-significant difference ($\gamma_{20} = 2.17, t = 1.42, p = .18$). Students in the NC + SRSD condition included approximately 2.7 more elements than students in the COMP condition in their essays at posttest ($\gamma_{30} = 2.66, t = 3.51, p < .01$). This model explained 18% of the within-teacher variability in posttest essay elements. Results of all models are presented in Table 4.6.
Table 4.6. *Unstandardized Coefficients (and Standard Errors) of Multilevel Models of Essay Elements*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Unconditional model</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest essay elements, $\beta_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>8.35 (.66)**</td>
<td>6.70 (.62)**</td>
<td>6.60 (.62)**</td>
</tr>
<tr>
<td>Pretest essay elements, $\beta_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{10}$</td>
<td>.54 (.04)**</td>
<td>.49 (.05)**</td>
<td></td>
</tr>
<tr>
<td>NC + TRAD condition, $\beta_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{20}$</td>
<td>2.11 (1.38)</td>
<td>2.17 (1.53)</td>
<td></td>
</tr>
<tr>
<td>NC + SRSD condition, $\beta_3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{30}$</td>
<td>2.68 (.81)**</td>
<td>2.66 (.76)**</td>
<td></td>
</tr>
<tr>
<td>Gender, $\beta_4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{40}$</td>
<td>.62 (.28)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous ELA achievement, $\beta_5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{50}$</td>
<td>.09 (.01)**</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Random Effects                        |                     |         |         |
| Posttest essay elements ($\tau_{00}$) | 7.84 (2.71)**       | 1.69 (.87)* | 1.24 (.69)** |
| NC + TRAD condition slope ($\tau_{11}$) | 7.27 (5.40)       | 10.18 (6.84) |
| Within-teacher variance ($\sigma^2$)  | 13.39 (.67)**      | 10.73 (.53)** | 11.02 (.63)** |
| Deviance (-2LL)                       | 4562.7              | 4366.5  | 3349.4  |
| Number of est. parameters             | 3                   | 7       | 9       |

*Note.* *p*<.05, **p**<.01, ***p**<.001

**Writing knowledge.**

**Planning knowledge.** For planning knowledge items, responses scored as substantive processes accounted for over 83% of all idea units, in-line with previous research (e.g., Harris et al., 2006). Production procedures, motivation, seeking assistance, environmental structuring, and ability accounted for only 2%, <1%, 7%, 3% and <1% of all pre- and
posttest planning knowledge responses, respectively; thus, these categories were not subjected to analysis.

Following the unconditional model for substantive processes knowledge, a second model (Model 2) was estimated to control for pretest knowledge of substantive processes and include treatment variables as predictors. Results showed that, averaging across students and within teachers, students identified more substantive processes at posttest than pretest ($\gamma_{10} = 0.26, t = 6.45, p < .001$). However, the number of substantive processes students identified at posttest did not vary significantly by condition ($\gamma_{20} = -0.31, t = -1.23, p = .24, \gamma_{30} = -0.01, t = -0.03, p = .98$). This model explained 20% of the within-teacher variability in posttest substantive processes knowledge.

Finally, successive third models (Model 3) were estimated to further explain the random variation in posttest substantive processes knowledge by adding controls for demographic variables and previous ELA achievement. Age, gender, race/ethnicity, LEP, and SWD status were found to be nonsignificant predictors of posttest substantive processes knowledge and were removed from the final model for parsimony. Results of the final model showed, averaging across students, within teachers, and controlling for student demographics and pretest substantive processes knowledge, students’ substantive processes knowledge at posttest did not vary significantly by condition ($\gamma_{20} = -0.32, t = -1.24, p = .23, \gamma_{30} = -0.01, t = .25, p = .96$). This model explained 20% of the within-teacher variability in posttest substantive processes knowledge. Results of all models are presented in Table 4.7.
Table 4.7.  
*Unstandardized Coefficients (and Standard Errors) of Multilevel Models of Substantive Processes Knowledge*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Unconditional model</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest substantive processes knowledge, $\beta_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>2.26 (.12)***</td>
<td>2.35 (.19)***</td>
<td>2.37 (.20)***</td>
</tr>
<tr>
<td>Pretest substantive processes knowledge, $\beta_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{10}$</td>
<td>.26 (.04)***</td>
<td>.27 (.05)***</td>
<td></td>
</tr>
<tr>
<td>NC + TRAD condition, $\beta_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{20}$</td>
<td>−.31 (.25)</td>
<td>−.32 (.26)</td>
<td></td>
</tr>
<tr>
<td>NC + SRSD condition, $\beta_3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{30}$</td>
<td>−.01 (.24)</td>
<td>−.01 (.25)</td>
<td></td>
</tr>
<tr>
<td>Previous ELA achievement, $\beta_4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{40}$</td>
<td>.01 (.005)*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Random Effects                          |                     |         |         |
| Posttest substantive processes knowledge ($\tau_{00}$) | .22 (.09)**        | .14 (.07)* | .12 (.07)* |
| Within-teacher variance ($\sigma^2$)    | 1.37 (.07)***       | 1.28 (.07)*** | 1.28 (.08)*** |
| Deviance (-2LL)                         | 2475.5              | 2258.0  | 1734.9  |
| Number of est. parameters               | 3                   | 6       | 7       |

**Argumentative writing knowledge.** For argumentative writing knowledge items, responses scored as elements accounted for over 48% of all idea units. Production procedures accounted for only 6% of all pre- and posttest argumentative writing knowledge responses and therefore was not subjected to analysis.

Following the unconditional model for essay element knowledge, a second model (Model 2) was estimated to control for pretest element knowledge and include treatment variables as predictors. Results showed that, averaging across students and within teachers, students identified more elements at posttest than pretest ($\gamma_{10} = 0.36$, $t = 9.54$, $p < .001$).
However, there were no significant differences in the number of elements identified by students in the NC + TRAD and COMP conditions at posttest ($\gamma_{20} = .12, t = .56, p = .58$). Students in the NC + SRSD condition identified significantly more basic elements of argumentative essays than students in the other two conditions at posttest ($\gamma_{30} = .52, t = 2.59, p < .05$). This model explained 11% of the within-teacher variability in posttest essay element knowledge.

Finally, successive third models (Model 3) were estimated to further explain the random variation in posttest essay element knowledge by adding controls for demographic variables and previous ELA achievement. Age, race/ethnicity, LEP, and SWD status were found to be nonsignificant predictors of posttest element knowledge and were removed from the final model for parsimony. Results of the final model showed, averaging across students, within teachers, and controlling for student demographics and pretest argumentative writing knowledge of essay elements, there were no significant differences between the number of elements identified by students in the NC + TRAD condition and students in the COMP conditions at posttest ($\gamma_{20} = .04, t = .18, p = .86$). However, students in the NC + SRSD condition identified approximately 0.6 more basic elements of argumentative essays than students in the COMP condition at posttest ($\gamma_{30} = .58, t = 2.56, p < .05$). This model explained 9% of the within-teacher variability in posttest essay element knowledge. Results of all models are presented in Table 4.8.
Table 4.8.  
*Unstandardized Coefficients (and Standard Errors) of Multilevel Models of Essay Element Knowledge*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Unconditional model</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest element knowledge, $\beta_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>1.50 (.12)***</td>
<td>1.22 (.15)***</td>
<td>1.10 (.19)***</td>
</tr>
<tr>
<td>Pretest element knowledge, $\beta_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{10}$</td>
<td>.36 (.04)***</td>
<td>.24 (.05)***</td>
<td></td>
</tr>
<tr>
<td>NC + TRAD condition, $\beta_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{20}$</td>
<td>.12 (.21)</td>
<td>.04 (.24)</td>
<td></td>
</tr>
<tr>
<td>NC + SRSD condition, $\beta_3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{30}$</td>
<td>.52 (.20)*</td>
<td>.58 (.23)*</td>
<td></td>
</tr>
<tr>
<td>Gender, $\beta_4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{40}$</td>
<td>.23 (.10)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous ELA achievement, $\beta_5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{50}$</td>
<td>.03 (.01)***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest element knowledge ($\tau_{00}$)</td>
<td>.22 (.09)***</td>
<td>.08 (.04)*</td>
<td>.08 (.05)</td>
</tr>
<tr>
<td>Within-teacher variance ($\sigma^2$)</td>
<td>1.46 (.07)***</td>
<td>1.30 (.07)***</td>
<td>1.33 (.08)***</td>
</tr>
<tr>
<td>Deviance (-2LL)</td>
<td>2528.0</td>
<td>2263.7</td>
<td>1747.3</td>
</tr>
<tr>
<td>Number of est. parameters</td>
<td>3</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>

*Note.* *p* < .05, **p* < .01, ***p* < .001

**Writing self-efficacy.** Following the unconditional model for writing self-efficacy, a second model (Model 2) was estimated to control for pretest writing self-efficacy and include treatment variables as predictors. Results showed that on average, students had higher writing self-efficacy at posttest than pretest ($\gamma_{10} = 0.68$, $t = 18.31$, $p < .001$). However, students’ writing self-efficacy at posttest did not vary significantly by condition ($\gamma_{20} = 3.89$, $t = 1.45$, $p = .17$, $\gamma_{30} = .90$, $t = .35$, $p = .73$). This model explained 31% of the within-teacher variability in posttest writing self-efficacy.
Finally, successive third models (Model 3) were estimated to further explain the random variation in posttest writing self-efficacy by adding controls for demographic variables and previous ELA achievement. Age, gender, race/ethnicity, LEP, and SWD status were found to be nonsignificant predictors of posttest writing self-efficacy and were removed from the final model for parsimony. Results of the final model showed, averaging across students, within teachers, and controlling for student demographics and pretest writing self-efficacy, students’ writing self-efficacy at posttest did not vary significantly by condition ($\gamma_{20} = 3.62, t = 1.23, p = .24, \gamma_{30} = 1.15, t = .41, p = .69$). This model explained 33% of the within-teacher variability in posttest writing self-efficacy. Results of all models are presented in Table 4.9.
Table 4.9.
*Unstandardized Coefficients (and Standard Errors) of Multilevel Models of Writing Self-Efficacy*

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Unconditional model</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest writing self-efficacy, $\beta_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>55.05 (1.35)*****</td>
<td>54.15 (1.96)*****</td>
<td>54.47 (2.29)*****</td>
</tr>
<tr>
<td>Pretest writing self-efficacy, $\beta_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{10}$</td>
<td>0.68 (.04)*****</td>
<td>0.63 (.04)*****</td>
<td></td>
</tr>
<tr>
<td>NC + TRAD condition, $\beta_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{20}$</td>
<td>3.89 (2.68)</td>
<td>3.62 (2.94)</td>
<td></td>
</tr>
<tr>
<td>NC + SRSD condition, $\beta_3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{30}$</td>
<td>0.90 (2.61)</td>
<td>1.15 (2.81)</td>
<td></td>
</tr>
<tr>
<td>Previous ELA achievement, $\beta_4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{40}$</td>
<td>0.16 (.07)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest writing self-efficacy ($\tau_{00}$)</td>
<td>21.71 (12.02)*</td>
<td>11.65 (7.31)</td>
<td>9.64 (7.68)</td>
</tr>
<tr>
<td>Within-teacher variance ($\sigma^2$)</td>
<td>430.77 (22.21)*****</td>
<td>295.74 (15.79)*****</td>
<td>287.63 (17.61)*****</td>
</tr>
<tr>
<td>Deviance (-2LL)</td>
<td>6911.3</td>
<td>6181.6</td>
<td>4729.2</td>
</tr>
<tr>
<td>Number of est. parameters</td>
<td>3</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

*Note. *$p<.05$, **$p<.01$, ***$p<.001$*

**RQ2:** What is the shape of growth in students’ writing quality, essay length, and essay elements associated with NC Write + traditional writing instruction and NC Write + SRSD instruction?

To examine the shape of growth in students’ writing quality, essay length, and essay elements associated with NC Write + traditional writing instruction and NC Write + SRSD instruction, a series of three-level models were specified to incorporate all essays written by
students in the treatment conditions. During the course of the intervention these students each completed 5–11 essays ($M=6.31$, $SD=.92$), including the pretest, posttest, and collaborative essay. Preliminary review of individual students’ essay performance (i.e., writing quality, essay length, and essay elements) revealed both variability in outcomes at pretest and variability in change over time. Many students appeared to show curvilinear growth in outcomes over time by demonstrating fairly rapid improvement in performance following the start of treatment but slower growth towards the end of the intervention. Therefore, both linear and quadratic growth was tested using a polynomial time model comprised of the variables $Time$ and $Time^2$. The $Time$ variable is a count of each essay completed and is used to describe instantaneous rate of change, while the $Time^2$ variable is used to describe the deceleration in students’ growth curves. Both variables were centered such that 0 represented students’ pretest writing performance.

The following model was used to examine the shape of growth in students’ writing quality, essay length, and essay elements associated with NC Write + traditional writing instruction and NC Write + SRSD instruction:

Level 1: $Y_{tij} = \pi_{0ij} + \pi_{1ij}(Time_{tij}) + \pi_{2ij}(Time^2_{tij}) + \epsilon_{tij}$

Level 2: $\pi_{0ij} = \beta_{00j} + r_{0ij}$

$\pi_{1ij} = \beta_{10j} + r_{1ij}$

$\pi_{2ij} = \beta_{20j} + r_{2ij}$

Level 3: $\beta_{00j} = \gamma_{000} + u_{00j}$

$\beta_{10j} = \gamma_{100} + u_{10j}$

$\beta_{20j} = \gamma_{200} + u_{20j}$
In this model, \( i \) and \( j \) continue to represent students and teachers, respectively, but \( t \) is added to denote time at Level 1. Thus, at Level 1 \( Y_{ij} \) represents the outcome for (1) writing quality, (2) essay length, and (3) essay elements predicted for student \( i \) with teacher \( j \) at time \( t \), \( \pi_{0ij} \) represents the initial status of student \( i \) with teacher \( j \) for the respective predicted outcome, \( \pi_{1ij} \) represents the instantaneous rate of change for student \( ij \), \( \pi_{2ij} \) represents deceleration in rate of change for student \( ij \), and \( e_{ij} \) represents the variance in the predicted outcome for student \( ij \) at time \( t \).

At Level 2, students’ intercepts (\( \pi_{0ij} \)) and slopes (\( \pi_{1ij} - \pi_{2ij} \)) become the outcomes. Here, \( \beta_{00j} \) represents the mean initial status averaging across students and within teachers. The average rate of change and deceleration within teachers is represented by \( \beta_{10j} \) and \( \beta_{20j} \), respectively. The residuals \( r_{0ij} - r_{2ij} \) represent variance around the intercepts and slopes for students within teachers.

Finally, at Level 3, the teacher-specific student intercepts (\( \beta_{00j} \)) and slopes (\( \beta_{10j} - \beta_{20j} \)) become the outcomes. Here, \( \gamma_{000} \) represents the mean initial status across all students and teachers, \( \gamma_{100} \) and \( \gamma_{200} \) represent the mean rate of change and deceleration across all students and teachers, and the residuals \( u_{00j} - u_{20j} \) represent variance around the intercepts and slopes for students across teachers.

An unconditional model (Model 1) was first specified to calculate intra-class correlations and ensure there was sufficient variance at all levels to examine a three-level model. Results of all unconditional models indicated there was sufficient variability at all levels to justify a three-level model (see Tables 4.10–4.12).
Writing quality. Following the unconditional model for writing quality, unconditional linear (Model 2) and quadratic (Model 3) growth models were estimated to examine the effect of time on writing quality. Due to the variance in rate of change and deceleration in writing quality over time both between and within students, allowing the slopes to vary randomly at Levels 2 and 3 provided better model fit than constraining the slopes, (Model 3 $\chi^2 (10) = 614.9, p < .001$). The quadratic model provided the best fit to the data and explained 41% of the within-student variability in writing quality. Results of this model showed that, averaging across students and within teachers, students improved the quality of their essays over time by approximately 1.6 points per essay ($\gamma_{100} = 1.55, t = 4.73, p < .001$) with a deceleration of −0.2 points multiplied by the respective values for each time variable ($\gamma_{200} = -0.21, t = -4.52, p < .001$). These parameters were used to estimate the saturation point at which growth slowed to zero. On average, students’ improvement in writing quality reached this saturation point after approximately the fourth essay written during the intervention.

To determine if students’ growth in writing quality varied by condition, a conditional model (Model 4) was estimated that added a condition variable as a predictor and included condition × time interactions. Results of this model showed nonsignificant differences in the rate of change ($\gamma_{101} = .66, t = 1.08, p = .28$) and deceleration ($\gamma_{201} = -.06, t = -.77, p = .44$) in writing quality over time for students in the NC + SRSD condition compared to students in the NC + TRAD condition.

Finally, successive models were specified that added Level 2 variables to further explain the random variation in both pretest writing quality and change in quality over time.
Level 2 variables included gender, previous ELA achievement, LEP, SWD, race/ethnicity, age, pretest writing self-efficacy, achievement goal orientation, and intervention fidelity. All continuous variables were grand-mean centered. For parsimony, all non-significant variables were excluded from the model. Gender, previous ELA achievement, Hispanic ethnicity, pretest writing self-efficacy, performance approach goal orientation, and intervention fidelity all predicted between-student variation in pretest writing quality but not change in quality over time.

After controlling for the demographic variables and previous ELA achievement, results of the final conditional model (Model 5) showed that compared to students in the NC + TRAD condition, on average students in the NC + SRSD condition improved the quality of their essays over time by an additional .5 points per essay ($\gamma_{101} = .49$, $t = .74$, $p = .46$) with a deceleration of $-.05$ points multiplied by the respective values for each time variable ($\gamma_{201} = -.05$, $t = -.45$, $p = .65$), though the differences between the two conditions were not significant. The final conditional model explained 61% of the between-student variability in writing quality. All models are displayed in Table 4.10.
Table 4.10.
Unstandardized Coefficients (and Standard Errors) of Multilevel Growth Models of Writing Quality

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing quality, $\pi_{0ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial status, $\gamma_{000}$</td>
<td>14.11 (.57)***</td>
<td>12.85 (.67)***</td>
<td>12.05 (.68)***</td>
<td>13.47 (.90)***</td>
<td>12.59 (.77)***</td>
</tr>
<tr>
<td>NC + SRSD, $\gamma_{001}$</td>
<td></td>
<td></td>
<td></td>
<td>$-2.51$ (1.21)</td>
<td>$-2.00$ (1.01)</td>
</tr>
<tr>
<td>Gender, $\gamma_{002}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.30 (.18)***</td>
</tr>
<tr>
<td>Race, Hispanic, $\gamma_{003}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.68 (.26)**</td>
</tr>
<tr>
<td>ELA achievement$^a$, $\gamma_{004}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.13 (.01)***</td>
</tr>
<tr>
<td>Writing SE$^b$, $\gamma_{005}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.01 (.005)*</td>
</tr>
<tr>
<td>Performance apr.$^c$, $\gamma_{006}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.08 (.03)**</td>
</tr>
<tr>
<td>Interven. fidelity, $\gamma_{007}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.19 (.05)***</td>
</tr>
<tr>
<td>Rate of change, $\pi_{1ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{100}$</td>
<td>.43 (.15)***</td>
<td>1.55 (.34)***</td>
<td>1.17 (.48)***</td>
<td>1.38 (.50)***</td>
<td></td>
</tr>
<tr>
<td>NC + SRSD, $\gamma_{101}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceleration, $\pi_{2ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{200}$</td>
<td>$-0.21$ (.05)***</td>
<td>$-0.17$ (.07)***</td>
<td>$-0.20$ (.08)***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC + SRSD, $\gamma_{201}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$-0.07$ (.09)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$-0.05$ (.10)</td>
</tr>
</tbody>
</table>
Table 4.10. Continued

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Level 1 (time)</th>
<th>Level 2 (students)</th>
<th>Level 3 (teachers)</th>
<th>Deviance (-2LL)</th>
<th>Number of est. parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.68***</td>
<td>6.42***</td>
<td>5.12***</td>
<td>5.12***</td>
<td>4.97***</td>
</tr>
<tr>
<td></td>
<td>(.25)***</td>
<td>(.18)***</td>
<td>(.18)***</td>
<td>(.18)***</td>
<td>(.21)***</td>
</tr>
<tr>
<td>Level 2 (students)</td>
<td>Initial status</td>
<td>4.86***</td>
<td>3.86***</td>
<td>4.33**</td>
<td>4.32***</td>
</tr>
<tr>
<td></td>
<td>(.39)***</td>
<td>(.48)***</td>
<td>(.57)***</td>
<td>(.57)***</td>
<td>(.48)***</td>
</tr>
<tr>
<td>Rate of change</td>
<td>- .22</td>
<td>.22</td>
<td>.62</td>
<td>(.17)</td>
<td>(.17)</td>
</tr>
<tr>
<td></td>
<td>(.17)</td>
<td>(.17)</td>
<td>(.17)</td>
<td>(.17)</td>
<td>(.17)</td>
</tr>
<tr>
<td></td>
<td>.003</td>
<td>.003</td>
<td>.008</td>
<td>.004</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td>(.004)</td>
<td>(.004)</td>
<td>(.004)</td>
<td>(.004)</td>
<td>(.004)</td>
</tr>
<tr>
<td>Level 3 (teachers)</td>
<td>Initial status</td>
<td>4.26**</td>
<td>5.93**</td>
<td>6.12**</td>
<td>4.68**</td>
</tr>
<tr>
<td></td>
<td>(1.74)**</td>
<td>(2.37)**</td>
<td>(2.46)**</td>
<td>(1.87)**</td>
<td>(1.31)**</td>
</tr>
<tr>
<td>Rate of change</td>
<td>.28</td>
<td>1.40</td>
<td>1.29</td>
<td>1.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.11)**</td>
<td>(.55)**</td>
<td>(.51)**</td>
<td>(.55)**</td>
<td></td>
</tr>
<tr>
<td>Deceleration</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.01)**</td>
<td>(.01)**</td>
<td>(.01)**</td>
<td>(.01)**</td>
<td></td>
</tr>
<tr>
<td>Deviance (-2LL)</td>
<td>15877.2</td>
<td>15169.1</td>
<td>14781.6</td>
<td>14777.7</td>
<td>11846.4</td>
</tr>
<tr>
<td>Number of est. parameters</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>13</td>
<td>19</td>
</tr>
</tbody>
</table>

Note. *p<.05, **p<.01, ***p<.001
aPrevious ELA Achievement based on scale score (range: 423–484) from the previous year’s end-of-grade assessment.
bWriting self-efficacy at pretest.
cPerformance approach goal orientation.

d**Essay length.** Following the unconditional model for essay length, unconditional linear (Model 2) and quadratic (Model 3) growth models were estimated to examine the effect of time on essay length. Due to variance in the rate of change in essay length over time between and within students, allowing the slopes to vary randomly at Levels 2 and 3 provided better model fit than constraining the slopes, (Model 3 $\chi^2 (10) = 692, p < .001$). The quadratic model provided the best fit to the data and explained 41% of the within-student variability in essay length. Averaging across students and within teachers, students increased the length of their essays over time by approximately 45 words per essay ($\gamma_{100} = 45.46, t = 4.13, p < .001$) with a deceleration of $-6$ words multiplied by the respective values for each
time variable ($\gamma_{200} = -5.96, t = -3.66, p < .001$). On average, students’ growth in essay length reached a saturation point after approximately the fourth essay written during the intervention.

To determine if students’ growth in essay length varied by condition, a conditional model (Model 4) was estimated that added a condition variable as a predictor and included condition $\times$ time interactions. Results of this model showed nonsignificant differences in the rate of change ($\gamma_{101} = 18.37, t = .85, p = .39$) and deceleration ($\gamma_{201} = -1.90, t = -.59, p = .57$) in essay length over time for students in the NC + SRSD condition compared to students in the NC + TRAD condition.

Finally, successive models were specified that added Level 2 variables to further explain the random variation in pretest essay length and change in length over time. In the final model the deceleration slope was constrained at Level 2 as there was insignificant variance around this slope and in order for the model to converge. Gender, previous ELA achievement, Hispanic ethnicity, pretest writing self-efficacy, and intervention fidelity all predicted between-student variation in pretest essay length but not change in length over time.

Results of the final conditional model (Model 5) showed that, averaging across students, within teachers, and controlling for demographic variables, compared to students in the NC + TRAD condition, on average students in the NC + SRSD condition increased the length of their essays over time by an additional 10 words per essay ($\gamma_{101} = 10.46, t = .45, p = .65$) with a deceleration of $-0.8$ words multiplied by the respective values for each time variable ($\gamma_{201} = -0.76, t = -0.20, p = .84$), though these differences between the two conditions were not significant. The final conditional model explained 56% of the between-student variability in essay length. All models are displayed in Table 4.11.
Table 4.11.  
Unstandardized Coefficients (and Standard Errors) of Multilevel Growth Models of Essay Length

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, $\gamma_{000}$</td>
<td>267.09 (22.30)***</td>
<td>227.63 (27.16)***</td>
<td>204.42 (26.46)***</td>
<td>263.57 (34.34)***</td>
<td>229.88 (31.70)***</td>
</tr>
<tr>
<td>NC + SRSD, $\gamma_{001}$</td>
<td></td>
<td></td>
<td></td>
<td>263.57 (34.34)***</td>
<td>229.88 (31.70)***</td>
</tr>
<tr>
<td>Gender, $\gamma_{002}$</td>
<td>52.06 (7.52)***</td>
<td>32.76 (10.96)**</td>
<td>52.06 (7.52)***</td>
<td>32.76 (10.96)**</td>
<td>52.06 (7.52)***</td>
</tr>
<tr>
<td>Race, Hispanic, $\gamma_{003}$</td>
<td>2.90 (.41)***</td>
<td>.52 (.23)*</td>
<td>2.90 (.41)***</td>
<td>.52 (.23)*</td>
<td>2.90 (.41)***</td>
</tr>
<tr>
<td>ELA achievement*, $\gamma_{004}$</td>
<td>2.22 (1.16)</td>
<td>5.22 (2.10)*</td>
<td>2.22 (1.16)</td>
<td>5.22 (2.10)*</td>
<td>2.22 (1.16)</td>
</tr>
<tr>
<td>Writing SEb, $\gamma_{005}$</td>
<td></td>
<td></td>
<td></td>
<td>5.22 (2.10)*</td>
<td>2.22 (1.16)</td>
</tr>
<tr>
<td>Performance apr.c, $\gamma_{006}$</td>
<td></td>
<td></td>
<td></td>
<td>5.22 (2.10)*</td>
<td>2.22 (1.16)</td>
</tr>
<tr>
<td>Interven. fidelity, $\gamma_{007}$</td>
<td></td>
<td></td>
<td></td>
<td>5.22 (2.10)*</td>
<td>2.22 (1.16)</td>
</tr>
</tbody>
</table>

Rate of change, $\pi_{1ij}$

| Intercept, $\gamma_{100}$ | 13.37 (5.86)* | 45.46 (11.00)*** | 35.09 (16.09)* | 18.37 (21.57) | 10.46 (23.21) |
| NC + SRSD, $\gamma_{101}$ |                          |                                     |                                         | 18.37 (21.57) | 10.46 (23.21) |

Deceleration, $\pi_{2ij}$

| Intercept, $\gamma_{200}$ | -5.96 (1.63)*** | -4.89 (2.41)* | -6.53 (2.81)* | -1.90 (3.23) | -7.6 (3.74) |
| NC + SRSD, $\gamma_{201}$ |                          |                                     |                                         | 1.90 (3.23) | -2.6 (3.84) |
Table 4.11. Continued

<table>
<thead>
<tr>
<th></th>
<th>Level 1 (time)</th>
<th>Level 2 (students)</th>
<th>Level 3 (teachers)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11998.00</td>
<td>8728.36</td>
<td>6749.62</td>
</tr>
<tr>
<td></td>
<td>(342.00)***</td>
<td>(281.71)***</td>
<td>(545.22)***</td>
</tr>
<tr>
<td></td>
<td>8728.36</td>
<td>3651.02</td>
<td>4363.25</td>
</tr>
<tr>
<td></td>
<td>(281.71)***</td>
<td>(665.86)***</td>
<td>(643.10)***</td>
</tr>
<tr>
<td></td>
<td>7033.21</td>
<td>794.87</td>
<td>22.86</td>
</tr>
<tr>
<td></td>
<td>(281.71)***</td>
<td>(643.10)***</td>
<td>(32.38)</td>
</tr>
<tr>
<td></td>
<td>7032.59</td>
<td>798.23</td>
<td>19.29</td>
</tr>
<tr>
<td></td>
<td>(273.85)***</td>
<td>(334.73)**</td>
<td>(9.73)*</td>
</tr>
<tr>
<td></td>
<td>7711.68</td>
<td>2945.07</td>
<td>19.37</td>
</tr>
<tr>
<td></td>
<td>(275.35)***</td>
<td>(37.78)**</td>
<td>(9.72)*</td>
</tr>
<tr>
<td>Level 2 (students)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial status</td>
<td>6749.62</td>
<td>4363.25</td>
<td>6607.56</td>
</tr>
<tr>
<td></td>
<td>(545.22)***</td>
<td>(643.10)***</td>
<td>(2663.60)**</td>
</tr>
<tr>
<td>Rate of change</td>
<td>22.86</td>
<td>794.87</td>
<td>460.50</td>
</tr>
<tr>
<td></td>
<td>(32.38)</td>
<td>(334.73)**</td>
<td>(179.36)**</td>
</tr>
<tr>
<td>Deceleration</td>
<td>19.29</td>
<td>798.23</td>
<td>32.09</td>
</tr>
<tr>
<td></td>
<td>(9.73)*</td>
<td>(334.46)**</td>
<td>(13.45)**</td>
</tr>
<tr>
<td>Level 3 (teachers)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial status</td>
<td>6607.56</td>
<td>9927.98</td>
<td>9297.98</td>
</tr>
<tr>
<td></td>
<td>(2663.60)**</td>
<td>(3889.54)**</td>
<td>(3683.37)**</td>
</tr>
<tr>
<td>Rate of change</td>
<td>460.50</td>
<td>1523.09</td>
<td>9398.55</td>
</tr>
<tr>
<td></td>
<td>(179.36)**</td>
<td>(521.22)**</td>
<td>(3683.37)**</td>
</tr>
<tr>
<td>Deceleration</td>
<td>32.09</td>
<td>1438.77</td>
<td>31.91</td>
</tr>
<tr>
<td></td>
<td>(13.45)**</td>
<td>(592.96)**</td>
<td>(13.14)**</td>
</tr>
</tbody>
</table>

| Deviance (-2LL)      | 37646.1         | 36898.5            | 36584.4            |
| Number of est. params| 4               | 7                  | 10                 |

Note. *p<.05, **p<.01, ***p<.001

*Previous ELA Achievement based on scale score (range: 423–484) from the previous year’s end-of-grade assessment.

*Writing self-efficacy at pretest.

*Performance approach goal orientation.

**Essay elements.** Following the unconditional model for essay elements, unconditional linear (Model 2) and quadratic (Model 3) growth models were estimated to examine the effect of time on essay elements. Due to the variance in rate of change in essay elements over time both between and within students, allowing the slopes to vary randomly at Levels 2 and 3 provided better model fit than constraining the slopes, (Model 3 χ² (2) = 44.2, p < .001). The deceleration slope was constrained at Level 2 as there was insignificant variance around this slope. The quadratic model provided the best fit to the data and explained 25% of the within-student variability in essay elements. Averaging across students and within teachers, students increased the number of elements in their essays over time by
approximately 1.6 elements per essay ($\gamma_{100} = 1.64, t = 4.80, p < .001$) with a deceleration of −.2 points multiplied by the respective values for each time variable ($\gamma_{200} = −.24, t = −4.42, p < .001$). On average, students’ growth in essay elements reached a saturation point after approximately the fourth essay written during the intervention.

To determine if students’ growth in essay elements varied by condition, a conditional model (Model 4) was estimated that added a condition variable as a predictor and included condition × time interactions. The deceleration slope was constrained at Level 2 based on evidence that there was insignificant variance around this slope and in order for the model to converge. Results of this model showed nonsignificant differences in the rate of change ($\gamma_{101} = .55, t = .82, p = .41$) and deceleration ($\gamma_{201} = −.06, t = −.59, p = .55$) in essay elements over time for students in the NC + SRSD condition compared to students in the NC + TRAD condition.

Finally, successive models were specified that added Level 2 variables to further explain the random variation in both pretest essay elements and change in elements over time. For parsimony, all non-significant variables were excluded from the model. Gender, previous ELA achievement, and pretest writing self-efficacy all predicted between-student variation in pretest essay elements but not change in elements over time.

Results of the final conditional model (Model 5) showed that, averaging across students and within teachers, and controlling for demographic variables, compared to students in the NC + TRAD condition, on average students in the NC + SRSD condition increased the number of elements in their essays over time by an additional .5 elements per essay ($\gamma_{101} = .51, t = .77, p = .45$) with a deceleration of −.07 elements multiplied by the
respective values for each time variable ($\gamma_{201} = -.07, t = -.60, p = .55$), though these
differences between the two conditions were not significant. The final conditional model
explained 80% of the between-student variability in essay elements. All models are displayed
in Table 4.12.

Table 4.12.
Unstandardized Coefficients (and Standard Errors) of Multilevel Growth Models of Essay
Elements

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial status, $\pi_{0ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{000}$</td>
<td>9.28 (.58)***</td>
<td>8.16 (.69)***</td>
<td>7.27 (.72)***</td>
<td>8.85 (.94)***</td>
<td>8.08 (.85)***</td>
</tr>
<tr>
<td>NC + SRSD, $\gamma_{001}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender, $\gamma_{002}$</td>
<td>1.05 (.23)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA achievement$^a$, $\gamma_{003}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing SE$^b$, $\gamma_{004}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate of change, $\pi_{1ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{100}$</td>
<td>.38 (.14)***</td>
<td>1.64 (.34)***</td>
<td>1.33 (.50)**</td>
<td>1.45 (.50)**</td>
<td></td>
</tr>
<tr>
<td>NC + SRSD, $\gamma_{101}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deceleration, $\pi_{2ij}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{200}$</td>
<td>-.24 (.05)***</td>
<td>-.20 (.08)*</td>
<td>-.22 (.08)**</td>
<td>- .22 (.08)**</td>
<td></td>
</tr>
<tr>
<td>NC + SRSD, $\gamma_{201}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.12. Continued

<table>
<thead>
<tr>
<th></th>
<th>Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
</tr>
<tr>
<td>Level 1 (time)</td>
<td>11.83 $(.34)^{***}$</td>
</tr>
<tr>
<td>Level 2 (students)</td>
<td></td>
</tr>
<tr>
<td>Initial status</td>
<td>6.29 $(.52)^{***}$</td>
</tr>
<tr>
<td>Rate of change</td>
<td>-</td>
</tr>
<tr>
<td>Deceleration</td>
<td>-</td>
</tr>
<tr>
<td>Level 3 (teachers)</td>
<td>4.36 $(1.81)^{**}$</td>
</tr>
<tr>
<td>Initial status</td>
<td>Rate of change</td>
</tr>
<tr>
<td>Deceleration</td>
<td>.26 $(.11)^{**}$</td>
</tr>
<tr>
<td>Deviance (-2LL)</td>
<td>16775.7</td>
</tr>
<tr>
<td>Number of est. params</td>
<td>4</td>
</tr>
</tbody>
</table>

Note. *p<.05, **p<.01, ***p<.001

aPrevious ELA Achievement based on scale score (range: 423–484) from the previous year’s end-of-grade assessment.
bWriting self-efficacy at pretest.

**RQ3: What are Students’ and Teachers’ Perceptions of the Events, Value, and Strengths and Weaknesses Associated with NC Write?**

**Survey results.** The first step in examining students’ and teachers’ perceptions of the events, value, and strengths and weaknesses associated with NC Write involved aggregating students’ and teachers’ responses from the post-intervention surveys. Table 4.13 summarizes students’ perceptions of NC Write at posttest based on survey responses. Table 4.14 summarizes teachers’ perceptions of NC Write following the intervention. Table 4.15 summarizes teachers’ perceptions of how much NC Write assists or impedes the writing development of particular groups of students.
Table 4.13.

*Student perceptions of NC Write at posttest*

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I find NC Write easy to use.</td>
<td>2.7</td>
<td>6.6</td>
<td>21.3</td>
<td>48.4</td>
<td>20.9</td>
</tr>
<tr>
<td>I sometimes have trouble knowing how to use NC Write.</td>
<td>26.9</td>
<td>37.4</td>
<td>18.8</td>
<td>15.1</td>
<td>1.7</td>
</tr>
<tr>
<td>I like writing with NC Write.</td>
<td>9.9</td>
<td>12.2</td>
<td>30.2</td>
<td>30.0</td>
<td>17.6</td>
</tr>
<tr>
<td>I revise my writing more when I use NC Write.</td>
<td>5.4</td>
<td>10.5</td>
<td>28.7</td>
<td>35.1</td>
<td>20.3</td>
</tr>
<tr>
<td>Writing with NC Write has increased my confidence in my writing.</td>
<td>5.6</td>
<td>7.9</td>
<td>24.6</td>
<td>33.7</td>
<td>28.1</td>
</tr>
<tr>
<td>NC Write provides good suggestions for improving my writing.</td>
<td>3.3</td>
<td>6.4</td>
<td>20.0</td>
<td>46.1</td>
<td>24.2</td>
</tr>
<tr>
<td>The essay scores NC Write gives are fair.</td>
<td>6.8</td>
<td>9.5</td>
<td>24.0</td>
<td>42.1</td>
<td>17.6</td>
</tr>
<tr>
<td>NC Write helps improve my writing.</td>
<td>3.9</td>
<td>6.0</td>
<td>20.9</td>
<td>40.9</td>
<td>28.3</td>
</tr>
<tr>
<td>The lessons in NC Write are helpful.</td>
<td>4.8</td>
<td>5.6</td>
<td>21.1</td>
<td>41.7</td>
<td>26.7</td>
</tr>
</tbody>
</table>

*Note. N* = 516.
Table 4.14.
*Teacher perceptions of NC Write post-intervention*

<table>
<thead>
<tr>
<th>Perception</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC Write gives fair and accurate scores</td>
<td>0</td>
<td>21</td>
<td>29</td>
<td>29</td>
<td>21</td>
</tr>
<tr>
<td>NC Write makes writing instruction easier.</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>71</td>
<td>21</td>
</tr>
<tr>
<td>NC Write saves me time.</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>NC Write makes teaching more enjoyable.</td>
<td>0</td>
<td>0</td>
<td>43</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>NC Write lets me focus on higher concerns of writing instead of mechanics.</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>57</td>
<td>21</td>
</tr>
</tbody>
</table>

*Note.* N = 14.

Table 4.15.
*Teacher perceptions of how much NC Write assists or impedes students’ writing development*

<table>
<thead>
<tr>
<th>Group</th>
<th>Very negative</th>
<th>Slightly negative</th>
<th>Neither positive nor negative</th>
<th>Slightly positive</th>
<th>Very positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Language Learners</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>50</td>
<td>29</td>
</tr>
<tr>
<td>Special education</td>
<td>0</td>
<td>0</td>
<td>36</td>
<td>43</td>
<td>21</td>
</tr>
<tr>
<td>Gifted</td>
<td>0</td>
<td>14</td>
<td>29</td>
<td>14</td>
<td>43</td>
</tr>
<tr>
<td>At-risk</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>79</td>
<td>14</td>
</tr>
<tr>
<td>General students, no special needs</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>79</td>
<td>14</td>
</tr>
</tbody>
</table>

*Note.* N = 14.

**Interview results.**

**Social validity of NC Write.** When asked whether using NC Write has helped them to become a better writer, 83% (25/30) of students interviewed said yes. The most common
reasons given were that feedback supported improvements and that students saw evidence of growth in their scores over the course of the intervention. Three of the five students who expressed less certainty were from the NC + TRAD condition. Two of these students stated that NC Write “kind of” helped them to become better writers; of these one student like the compositional aspects and the feedback but felt the lessons should be more engaging and the other student liked the feedback but felt the lessons should be more interactive. The third student from the NC + TRAD condition indicated that NC Write helped “to a certain extent” but that the scores and feedback were not as accurate as those provided by a human. The remaining two students who were less certain were from the NC + SRSD condition. One student stated that NC Write “kind of” helped; this student liked using graphic organizers for planning, but seemed dismayed by lack of growth. This student appeared to use feedback solely to revise essays to address grammar and spelling mistakes, and was frustrated by the lack of improvement in score. The last student indicated that NC Write helped “a little bit,” citing growth in total essay score over time from 12 to 18 points.

When asked whether the writing intervention was appropriate for a variety of students, 100% (12/12) of teachers interviewed said yes. The most common reasons given were that the intervention met students where they were in terms of writing development and that the treatment was individualized and allowed students to work at their own pace. Two teachers provided caveats regarding the intervention’s appropriateness for all students: one shared that struggling writers needed significant support to interpret feedback, and the other indicated that having greater flexibility regarding lessons, prompts, and pacing for the intervention would have allowed for greater individualization.
Finally, when asked whether they would recommend the writing intervention to other teachers, 100% (12/12) of teachers interviewed said yes. The most common reasons involved the available resources for writing such as lessons, prompts, and portfolios, and the immediate feedback provided by NC Write. Teachers indicated the resources and automated feedback saved them time in planning, reading essays, and proving their own feedback, and allowed students to write more than they would have otherwise.

Social validity of SRSD procedures and methods. When asked whether they liked the STOP planning strategy, 87% (13/15) of students in the NC + SRSD condition who were interviewed stated yes. The most common reasons for liking the STOP strategy were that it was easy to remember, organizing ahead of time made writing easier, and it was important to evaluate evidence before taking a side. For example, a student shared, “It’s very specific; it’s not confusing about what it tells you to do.” Of the two students who did not say yes, one student felt the STOP strategy required too many steps and the other student indicated STOP was not as important as the DARE strategy, positing that writers should already know to take a side and plan more as they write.

The DARE strategy for argumentative essay elements was also viewed positively by students, with 87% (13/15) of students interviewed in the NC + SRSD condition indicating they liked the strategy. The most common reasons for liking the DARE strategy were it tells each step for writing an argumentative essay and it can be used to check back over an essay. A student described how he used DARE:

I find it like a list. It goes down to what I need first—develop my claim—so I know, OK, I need to get this in order and this in order, and the reasons and evidence is big
for me. I need to have a lot of evidence to support the claim and then rejecting
counter claims like why I don’t like it . . . it just really helps me out.

Of the two students who did not say yes, one student found the development of claims and
addition of supporting reasons and evidence too similar. The other student expressed
frustration about not knowing where to reject counterclaims in an essay.

Students were less positive about goal setting, with only 53% (8/15) of those
interviewed reporting that they liked setting goals. The most common reasons for liking
setting goals were that they helped students improve and they could be adjusted to become
more challenging. A student explained:

If my goal were to add a lot of evidence and reasons I’d feel great as I checked it off
because I’d know that’s a goal I achieved and I got a higher score because I looked at
the STOP and DARE strategy.

Of the seven students who did not indicate they liked setting goals, one student expressed
indifference, one student did not feel like setting goals, one student did not like the
responsibility of setting goals for each essay, one student felt the value of goal-setting
depended on purpose of the essay (i.e., felt goals were not important for an initial draft but
were important for a final essay), one student liked setting goals only when the prompt was
particularly interesting, and two students did not remember setting goals.

Finally, 93% (14/15) of students interviewed shared that they liked using self-
statements. The most common reasons for liking self-statements were they were good
reminders to check whether all essay parts were included and evaluate the quality of each
essay part and they could be used for self-encouragement. For example, a student described
how she used self-statements to begin writing. “I thought, let’s get started—that’s my main thing—let’s just calm down, let’s get started, think of great ideas and then you’ll be fine.”

One of the students interviewed reported not using self-statements.

When asked whether they would recommend the STOP and DARE strategies to other teachers, 83% (5/6) of teachers interviewed said yes. The most common reasons for this were the strategies were easy to understand, provided students with a plan for writing, and guided students though the writing process. A teacher explained:

[The students] were able to memorize STOP and DARE because we did it so often and they could tell me not only what each letter stands for but what it actually means. These are students that before we started this program, they didn’t know what a counterclaim was. They probably didn’t know what a claim was in that language.

The teacher who would not recommend the strategies indicated this was due to the current lack of emphasis on writing, even among other English teachers. She stated, “I don’t feel like they place enough value on [writing] to be receptive to the strategies.”

**A framework for deliberate writing practice.** Multi-level model results showed that, controlling for pretest performance, NC + SRSD students produced posttest essays that were of a higher quality, longer, and included more basic elements of argumentative essays than students in the other two conditions. NC + TRAD students produced higher-quality essays than COMP students at posttest. However, results of further multi-level models that incorporated all essays produced by students in the treatment conditions showed no significant differences between NC + TRAD and NC + SRSD students in rates of change and deceleration of writing quality, essay length, and essay elements.
Qualitative analyses revealed that NC Write provided a framework for deliberate writing practice. Concepts that emerged from interview data help to explicate how NC Write supported students’ growth in writing performance and offer some explanation as to how students in the NC + TRAD condition were able to demonstrate comparable growth to their peers in the NC + SRSD condition. The framework for deliberate writing practice is shown in Figure 4.1. 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 that supported many of the key requirements of deliberate practice. In particular, the learning, practice, and feedback components allowed students frequent and suitable practice opportunities and task-level performance feedback regarding the quality of their writing. Interview results provided some evidence that these components also supported students’ intrinsic motivation. The interconnected nature of the components worked in concert to support students’ growth in writing quality. Concepts that emerged from interview data and help to explain how NC Write supported students’ growth in writing performance are discussed below.
Writing quality feedback. The most common concept addressed in interviews related to the automated writing quality feedback students received for each essay composed and revised in NC Write. One property of this task feedback was that there was variation in how students used the feedback. Some students focused on the total essay score, comparing it to previous essay scores, peers’ scores, and/or goals they had established. While some students devoted their attention to the spelling and grammar feedback, other students reviewed the more detailed descriptive analysis and feedback before revising. It was common for students to have particular traits they wanted to improve in mind and focus on the relevant feedback. For example, one student explained, “I read everything, but I focus on the Sentence Structure a lot more than I do the others because that’s what I struggle with most of the time.” Some
students compared their performance across the trait scores to identify areas of focus, such as
the student who described, “I looked at my [trait] scores . . . to see what I need to fix then I
went down and if I didn’t have a three or higher on them I would go back and fix those.”
Some teachers felt that students honed in on the visual components of the feedback such as
the spelling and grammar annotations on their essays and the bar graph of trait scores.

In general, teachers liked that students received individualized task feedback for each
essay composed and revised. However, a second property of the writing quality feedback was
the importance of teacher assistance with its interpretation. One teacher described her role in
helping students make sense of feedback:

I feel like it’s my job to make sure that they understand what the feedback means.
You had to sit down and go over it with them . . . . They really understood feedback
on conventions; they understood spelling and grammar; they really understood what
they were supposed to do there. They understood the organization; they did pretty
well with that. It was just a few of the things.

For the most part teachers did not characterize their role in helping students interpret
feedback as a limitation of the feedback. For example, one teacher described how the
automated feedback prompted valuable conversation:

It’s kind of funny that that information caused them to come to me and discuss it with
me as, “Why is this?” or “Why is that?” and maybe if I had just, oh, made a comment
on their paper and made a suggestion, we might not have discussed that like we did.
So it ended up being a positive.
A third property of writing quality feedback was that there were specific ways in which it supported improvement. 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 also described how the automated feedback supported these improvements. Both the quality and quantity of feedback afforded improvements in students’ writing performance. One student explained, “It’s just more feedback than an actual teacher would give you, I believe.” Another student felt, “It’s more detailed and it explains it a lot better than the teachers can sometimes.” In their interviews students also indicated the specificity of feedback allowed them to improve their writing. One student explained, “It makes it very obvious what you did or didn’t do right.” Over time, students drew conclusions regarding connections between their essays and the automated scores and writing analysis. One student shared, “We figured that using a counterargument brings your score up a whole lot more.” A teacher elaborated on how the feedback encouraged critical examinations of students’ writing and the scoring criteria:

I felt like the identification of some of the mistakes was the most effective for us, and then the writing consistently and seeing those scores go up as they made their editing changes and being able to let kids compare their writing to other students saying, “OK, if you worked with this partner, can you look at what this kid did in their transitions and the content that they put in compared to yours?” and “Why did you get this score versus why did you not get the other scores?”
Some teachers also noted that receiving regular and immediate feedback for each essay and revision appeared to support students’ self-monitoring. One teacher in the NC + TRAD condition described how 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.

A final property of writing quality feedback was that there were several limitations of the automated feedback. Some students had difficulty interpreting the feedback on their own. This was largely due to the domain-specific language that comprised the feedback. One student recalled, “At first I didn’t know what half that stuff was.” Some students found the quantity of feedback overwhelming. These students typically focused on the spelling/grammar feedback and total essay and trait scores. One student explained, “I used the bar graphs a lot, but I didn’t really use [the writing analysis] because it was a lot of reading. I didn’t wanna read it.” Only one of the students interviewed expressed concerns regarding the scoring accuracy. “I feel like it gives you good feedback . . . but then I feel like it’s not as accurate as an actual person grading it.” When asked to elaborate the student acknowledged, “I don’t really trust computers as much as I do an actual person. I feel like if I had an actual person then it feels more real.”

Another limitation of the feedback was its generality. Some students indicated the writing analysis would have held greater utility had it been more specific. One 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 also observed a lack of variety in aspects of the writing analysis. These opinions were shared by teachers, who in general found the feedback accurate and relevant but not sufficiently specific to be reliably actionable. Some teachers described the feedback as generic and prescriptive. These concerns were most poignant as students’ writing quality growth neared a plateau. Consistent with the quantitative findings, teachers and students described a point where students reached a ceiling in writing quality and had difficulty improving further. One 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 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?’” Another teacher suggested augmenting the feedback to include models of exemplary writing:

When it gave information about development or . . . having a stronger conclusion, that kind of thing, I could see in their writing they needed a stronger conclusion, but . . . . I think it would be great if it could give some type of example to help the kids understand that piece . . . . I think that would be, from a teacher, very effective in helping kids become better writers on a very individualized basis. Boy, that’s the challenge, and that’s what a lot of these types of programs help us as teachers do. Another limitation of the feedback raised by a few students was that at times they
made revisions and were discouraged by the lack of change in scores or feedback. Given the visual nature of the spelling/grammar feedback some students addressed these issues only—believing that they corrected all problems—but producing only surface-level revisions. 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 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 most common criticism of the feedback in students’ interviews was that a spell check should have been available while writing. Students were accustomed to being able to check spelling using tools such as spell check and autocorrection in web browsers and spell check on mobile devices. Though NC Write was designed without a spell check to allow students to practice spelling and provide a more accurate measure of conventions accuracy this caused some frustration for students.

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 in writing practice opportunities. A teacher explained, “The more you write, the better you’re going to be, so it just gives them the opportunities that they need.” Another teacher indicated
how, compared to the number of writing opportunities she is traditionally able to provide, “It just allows them to write more, faster.” Teachers also pointed to the immediacy of feedback as valuable to students. One described:

As teachers, if we have 65 or 70 essays to grade, we’re going to be delayed and the longer you’re away from that topic, the less, I think, that the student will connect to it. They just move on. So, I like the promptness of it.

Teachers also described how automated feedback was efficient in saving them time. “I think that’s the number one selling point for me: that I didn’t have to sit there and read every single one to find all the grammar mistakes, the spelling mistakes, that kind of thing.” That NC Write included lessons, prompts, graphic organizers, and portfolios resulted in additional efficiencies. One 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 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.

Like teachers, students also described how they were able to write more with NC Write. An additional dimension of this efficiency present in student interviews was the difference
between writing longhand and typing. For example, one student explained, “When I’m on the computer, I have more time because I can type faster than writing.”

**Evidence of growth.** When asked whether NC Write helped them to become better writers, students often spoke to the fact the NC Write provided them with visible evidence of growth. Students were generally cognizant about quality of their writing and whether it had improved during the intervention. For example, one student explained how 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.” Students described using their portfolios to monitor growth, and liked that they had objective measures of their writing quality and growth. One 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 relations among essay elements, effort, and writing quality was emphasized by NC Write providing visible evidence of growth. The automated scores and feedback made explicit to students that essays with more essay elements and/or fewer errors tend to be of a higher quality. “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.”

**Supporting writing instruction.** Another concept that emerged from interview results was that NC Write supported classroom writing instruction. NC Write was a tool teachers used to reinforce and student used to apply strategies introduced in the classroom. For example, students in the NC + SRSD condition were introduced to the STOP and DARE
strategies in a classroom 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 in the NC + SRSD condition 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 something.”

A second property of NC Write supporting writing instruction was alignment with classroom writing instruction. Some teachers described how the program fit with their approach to writing instruction and the writing standards. One teacher from the NC + TRAD condition 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 also made adjustments to their instruction to increase this alignment. A 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 writing instruction and practice. For some teachers in the NC + TRAD condition who did not have a preferred approach to teaching writing, the program structure of lessons, composition and revision opportunities, and feedback was seen as attractive. One teacher acknowledged:
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 increased attention on writing in the classroom. In their interviews the majority of teachers indicated that writing in general received less consideration than in the past, in part due to the elimination of the NC Writing Assessment Program. One 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.”

Supporting intrinsic motivation. Student and teacher interviews provided some evidence that NC Write supported students’ intrinsic motivation, though no quantitative data was collected to confirm this. Interview data suggested that NC Write afforded three major sources of intrinsic motivation: challenge, curiosity, and control (Lepper & Hodell, 1989). One teacher explained how NC Write challenged students while prompting their curiosity:

Kids were challenged to try to get those higher scores throughout the time; I thought that that was a really big positive with that program. It helped my kids improve their writing because they wanted to try to figure out how to get certain scores, to get their better score the first [draft] and then to be able to see their score continue to go up as they made editing changes or the revisions they were making, so I think their writing definitely improved.
NC Write also appeared to support students’ control over their writing performance. As previously described, the automated scores and feedback provided students with visible evidence of their growth, which qualitative results suggested emphasized the relations among essay elements, effort, and writing quality to students. Also, students largely worked autonomously and at their own pace with NC Write. Providing students with a choice of prompt for each essay allowed another aspect of control. One teacher shared, “I thought the prompts were good; I liked giving the kids options between choosing one of the two prompts, I really liked that in the writing piece.” Another teacher explained how choice of prompt allowed students to choose a topic of which they had some background knowledge. “I don’t recall them saying ‘I don’t know about either of these.’ Usually, if one [was unfamiliar], then they would automatically just go to the other one. So it was good that they had a choice.”

In their interviews some students suggested the cycles of practice and feedback provided efficacy information that may have supported their intrinsic motivation, though quantitative results showed no significant changes in self-efficacy as a result of the intervention. One student from the NC + SRSD condition explained:

> It just makes me analyze a lot more things as I’m writing and there’s a lot of strategies that I know and that I use . . . and it just makes me feel great about it because I know that I’ve . . . accomplished a certain goal.

Observations of this nature were made by students in both treatment conditions. A student from the NC + TRAD condition shared:

> I think the part where you can revise it— it gave you that opportunity to make yourself better. And it gave me more confidence, and it made me feel I was a better
writer. And as I continue to go through, as scores continue to get better, I feel like it’s helped me really think about the way that I write and how I write throughout whole thing.

Finally, students’ situational interest in NC Write and/or individual interest in technology may also have supported their intrinsic motivation. For example, one student expressed, “I liked it because you didn’t just have to sit in class and write papers. You had to type. I like that it’s on the computer.” Another student indicated a preference for the format and control she associated with NC Write, disclosing, “I have a hard time listening during class . . . and on the computer, I can concentrate more because I learn it by myself.”

Limitations of the framework. Students’ growth in writing quality, essay length, and essay elements was best represented by a quadratic growth model. Interview results shed light on aspects of the intervention and NC Write that may have limited growth and/or explain the deceleration in growth over essays. The limitations of feedback were discussed above; the remainder of this section is devoted to other limitations raised in interviews.

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 seemed to like the interactive nature of the lessons. Some teachers also like the lessons and found them effective. One 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 drew greater criticism than other components of NC Write. The most common concerns raised by students was that the lessons should have
been less challenging and dense with information. Some students also felt the lessons should have been more interactive and entertaining. In general students did not view the interactions as checks for understanding and seemed to expect more individualized feedback, likely juxtaposing the lesson and essay feedback. One student explained:

You can’t really get into the lesson and learn from it really. You can learn from it, but you’re not getting asked questions to make sure that you understand it. And it just seems if it was more engaging to us, then we would be able to understand it and relate to it and be able to use it in our writing, too.

Teachers’ primary criticism of the lessons was that they did not provide sufficient data about student performance. One teacher explained:

They didn’t have really good feedback . . . . The kids could click through and finish and . . . I did not know how effective that application through that process worked because there wasn’t really an end piece for me.

Another teacher shared a similar concern, expressing, “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.” A different teacher explained that since each lesson was aligned with a specific writing trait, a lesson score or some other quantitative feedback about students’ performance would have been helpful for diagnosing specific strengths and weaknesses.

In their interviews students and teachers were asked what they would change about NC Write. Many of students’ recommendations dealt with features of NC Write were not utilized for the present study, such as peer review functionality. Both students and teachers
suggested that model essays exemplifying the higher end of the score point range would have been helpful. NC Write included annotated sample essays at the time of the intervention but these were limited in number and not clearly linked to total essay scores. Multiple teachers felt that NC Write needed to be able to scan for plagiarized or copied text. Plagiarism was a greater concern in the upper grades, though sixth grade teachers also reported some instances of students copying the prompt. One 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 during the intervention as unobserved instances resulted in missed writing practice opportunities for students, who also received potentially misleading feedback about their writing performance.
Overview of Findings

Process-based approaches to writing do not provide the amount of support and scaffolding many students need to develop into effective writers (Graham et al., 2005; Harris, Graham, & Mason, 2006). These approaches tend to overlook the self-regulatory skills and motivational beliefs required for proficient writing (Harris, Santangelo, & Graham, 2008). Additionally, restricted writing opportunities preclude the sustained deliberate practice needed to develop expertise in writing (Kellogg & Whiteford, 2009). To address these concerns the present study examined an intervention designed to leverage the strengths of the SRSD model with the writing practice opportunities and automated feedback afforded by an AWE program.

The present study investigated changes in students’ writing performance, knowledge, and self-efficacy associated with NC Write + SRSD instruction and NC Write + traditional writing instruction compared to traditional writing instruction. The shape of growth in students’ writing quality, essay length, and essay elements associated with NC Write + traditional writing instruction and NC Write + SRSD instruction was also evaluated. Finally, the present study employed surveys and interviews to examine students’ and teachers’ perceptions of the events, value, and strengths and weaknesses associated with NC Write.
Changes in students’ writing performance, knowledge, and self-efficacy. Results of multi-level models that controlled for pretest performance and predicted posttest performance across students and within teachers showed that students in the NC + SRSD condition produced posttest essays that were of a higher quality, longer, and included more basic elements of argumentative essays than students in the other two conditions. SRSD implementation in the current study was unique in that it was implemented by teachers and supported in NC Write, and these findings augment previous SRSD research which has documented effects of SRSD instruction on writing performance and knowledge (Graham 2006; Graham & Harris, 2003; Graham et al., 2012; Graham & Perin 2007a). Results also showed that students in the NC + TRAD condition produced higher-quality essays than COMP students at posttest. This finding adds empirical evidence to the limited body of AWE research by showing that teacher facilitation of NC Write had a significant, positive impact on students’ writing quality. The design and methodology of the present study overcame various limitations of previous AWE research, including irrelevant prompts that limited usage (Grimes & Warschauer, 2008), automated feedback that was used primarily to make surface-level revisions (Wilson & Andrada, 2013), and research designs that did not allow determination of causality (Roscoe & McNamara, 2013).

Students in the NC + SRSD condition did not demonstrate greater substantive processes knowledge than their peers at posttest as predicted. On average, students in all three conditions showed growth in substantive processes knowledge from pre- to posttest. One possible explanation for the lack of between-condition differences is that declarative and even procedural knowledge of the STOP planning strategy did not translate to declarative
knowledge of planning more generally for students in the NC + SRSD condition. Some previous SRSD research has found effects on writing quality but not planning knowledge of substantive processes (Graham et al., 2005), in line with the current findings. A second possible explanation is that using a survey rather than an interview to collect writing knowledge responses introduced construct-irrelevant variance. Students may have provided more restricted responses than they would have orally, and results may conflate students’ writing knowledge with writing ability or motivation.

Finally, students’ writing self-efficacy at posttest did not vary significantly by condition. This result aligns with some other SRSD findings (Graham et al., 2005; Limpo & Alves, 2013). Due to students’ limited writing and feedback opportunities, it is likely that students overestimated their writing self-efficacy at pretest largely due to an absence of information on which to base their efficacy judgments.

**Shape of growth in students’ writing quality, essay length, and essay elements.**

Results of multi-level models examining the shape of growth in writing performance across students and within teachers showed that treatment students’ growth in writing quality, essay length, and essay elements was best represented by a quadratic growth model. On average, students’ growth in writing performance reached a plateau following the fourth essay written during the intervention. Results are similar to Ennis et al. (2015), who documented growth in secondary students’ writing performance during the first five weeks of an eight week SRSD treatment. Ennis et al. (2015) did not include a comparison or other condition against which to compare SRSD students’ rate and shape of growth.
Results also showed non-significant differences in rates of change and deceleration in writing quality, essay length, and essay elements favoring the writing performance of students in the NC + SRSD condition. There are a number of possible explanations for the lack of significant differences between treatment conditions. It may be that differences in mean changes in writing performance from essay to essay were too minor, and standard errors too large, to observe between-condition differences. Another possible explanation is that the SRSD model was not implemented with sufficient fidelity to maximally impact students’ writing performance. This seems unlikely as on average, the intervention fidelity for students in the NC + SRSD condition was higher than that of their peers in the NC + TRAD condition. However, the bulk of the extant SRSD research has been implemented by one research team or by teachers monitored by this research team following extensive professional development. Additionally, treatment intensity of these interventions has been high, with lessons being implemented on average three to five days per week (Harris et al., 2008b). One purpose of the current study was to examine practical implementation of an SRSD intervention by teachers a few days a week and without extensive training, and while treatment intensity is a limitation of the present study results have high ecological validity. A third explanation is that despite differences in instruction between the two treatment conditions the extensive opportunities to learn, practice, and receive feedback afforded by NC Write moderated the effect of writing instruction on writing performance. The hypothesis is further interrogated below.

Though students’ gender, previous ELA achievement, and pretest writing self-efficacy consistently predicted their pretest writing performance, none of these variables
related to students’ growth in writing quality, length, or number of basic elements. This finding is similar to that of Wilson et al. (2014) regarding the impact of automated feedback on writing quality growth, and is similarly affirming in showing that despite initial individual differences, all groups of students demonstrated approximately equivalent growth in writing performance on an essay to essay basis during the intervention.

Students’ and teachers’ perceptions of the events, value, and strengths and weaknesses associated with NC Write. Survey results showed students held generally favorable opinions of NC Write. For example, the percentage of students who agreed or strongly agreed with the statements “NC Write helps improve my writing,” “NC Write provides good suggestions for improving my writing,” “The lessons in NC Write are helpful,” and “The essay scores NC Write gives are fair,” was 69.2%, 70.3%, 68.4%, and 59.7%, respectively. Teachers also held generally favorable opinions of the writing program; the percentage of teachers who agreed or strongly agreed with the statements “NC Write saves me time,” “NC Write makes writing instruction easier,” “NC Write lets me focus on higher concerns of writing instead of mechanics,” and “NC Write gives fair and accurate scores,” was 86%, 92%, 78%, and 50%, respectively. These results showed that teachers held more favorable opinions of NC Write as an AWE program than Grimes and Warschauer (2010) found teachers held of My Access. For example, only 63% of teachers agreed or strongly agreed that “My Access makes writing instruction easier,” and only 29% of teachers agreed or strongly agreed that “My Access gives fair and accurate scores.” These findings may reflect general improvements in AWE technology rather than between-program differences. The percentage of teachers who indicated NC Write either slightly or very positively assists
the writing development of ELL, special education, gifted, at-risk, and general students was 79%, 64%, 57%, 93%, and 93%, respectively. Grimes and Warschauer (2010) reported a higher percentage of teachers holding positive attitudes for My Access assisting the writing development of special education and gifted students, at 81% and 67%, respectively. Across studies teachers have held less positive attitudes towards AWEs assisting the writing development of gifted students; this area would benefit from additional research.

Interview results determined that NC Write as well as the greater writing intervention had acceptable social validity. The majority of students interviewed felt that using NC Write helped them to become better writers, substantiating the survey results. All teachers interviewed felt that the writing intervention was appropriate for a variety of students and indicated they would recommend the intervention to other teachers. Regarding the SRSD procedures and methods specifically, the vast majority of students interviewed liked the STOP and DARE strategies and using self-statements. Students were less positive about goal setting, with slightly over half the students interviewed reporting liking setting goals. Only one of the SRSD teachers interviewed indicated she would not recommend the STOP and DARE strategies to other teachers and this was because she felt her colleagues placed little value on writing instruction.

Qualitative data analysis revealed that NC Write provided a framework for deliberate writing practice. In this framework, students’ growth in writing performance is supported by a cycle of learning, practice, and feedback. Specifically, NC Write enabled deliberate practice throughout this cycle by affording writing quality feedback, efficiency, and evidence of growth, and supporting teachers’ writing instruction and students’ intrinsic motivation.
Limitations of the framework included some aspects of feedback, limited lesson data, and a lack of a plagiarism scanner in NC Write.

The efficiency of NC Write allowed students more writing opportunities than teachers could support without technology. NC Write also provided students with more feedback than teachers alone could provide. Though students required some assistance to interpret this feedback they were generally able to employ it in essay revisions and subsequent compositions and observe a commensurate improvement in writing quality. Collectively, the cycles of learning, practice, and feedback may have helped demystify the construct of writing quality for students. To provide another example, when students leaned a new strategy (from a lesson in NC Write or from their classroom teacher), they were able to implement the strategy immediately and observe the impact on their writing quality. Interview results suggested that over time students examined the relations among essay elements and the writing quality scores assigned by PEG, and in doing, likely developed a better understanding of the components of good argumentative writing. Experience alone is insufficient for deliberate practice (Ericsson, 2006), and interview results suggested the interconnected nature of the learning, practice, and feedback cycles of NC Write explicated both current levels of performance and components necessary for improved performance. This information allowed students to concentrate on improving specific aspects of performance and not just writing more.

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). Interview results suggested that the automated
feedback supported students’ self-monitoring and self-evaluation of their writing. That automated scores and feedback were immediately available and, on average, perceived as valid and valuable may have fostered students’ self-evaluation. This, in turn, likely helped students see the relations between effort and writing performance. One hallmark of SRSD is the teaching of self-regulatory strategies. Though NC + TRAD students were not explicitly taught self-regulatory skills, interview results provided moderate evidence that utilizing a framework for deliberate writing practice 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. This may explain why the writing performance growth of NC + TRAD students was not significantly different from that of NC + SRSD students on an essay to essay basis.

Interview data regarding feedback substantiated survey results, showing that teachers and students held generally positive opinions about the automated feedback and that there were various ways the feedback supported students’ improvement in writing performance. However, several limitations of feedback were apparent. Students had difficulty interpreting the feedback due to the language used. Writing expertise requires both general strategy and domain-specific knowledge (Kellog & Whiteford, 2009), and revising feedback language to be more interpretable by a novice would likely be detrimental in the long term. A more appropriate solution would be to add pop-up definitions and/or examples to the domain-specific language that comprises the feedback to support students’ development of domain knowledge and reduce the amount of teacher assistance required. Students and teachers criticized the lack of feedback specificity. General feedback introduced ambiguity regarding
how to improve essays and subjectivity in interpretation. Roscoe and McNamara (2013) reported the same concerns from an investigation of W-Pal, and described the development of more individualized feedback as the greatest challenge for AWE programs. AWE program feedback is inherently limited to those features of text that can be reliably measured (Roscoe & McNamara, 2013); in the case of PEG the challenge lies in translating unobservable features extracted from the training data into actionable feedback. This is a daunting challenge as most of the features that PEG identifies and extracts automatically during model building are unobservable or minimally impractical for humans to calculate.

Intervention timing may not have allowed students to produce their best writing at all timepoints. All but one teacher administered posttest essays during the last month of school. In their interviews teachers noted an observable decrease in students’ effort at the end of the year, and teachers generally felt that students’ writing quality was depressed towards the end of the intervention as a result of the timing.

Interview results supported a number of conclusions to enable students to maintain writing quality growth rates. First, automated feedback should be as specific and individualized as technology allows. If students’ essay revisions are insufficient to impact their writing quality some information to this effect should be provided by the program. Given the complexity of the scoring and feedback models, even minor revisions provide information about a student’s command of the writing construct that could be used to adjust feedback, even if this is just to acknowledge the revision. Second, automated feedback needs to be more accessible and actionable. This may require supports to help students interpret domain-specific language and (re)structuring feedback in such a way that makes clear to
students what steps to take to improve writing quality. The latter would involve providing examples, non-examples, and model essays to explicate various aspects of writing quality. Third, students should be provided with carefully designed practice opportunities that can be mastered sequentially. This would have the effect of differentiating deliberate practice from writing experience. Such practice opportunities should allow students to apply specific skills and strategies in their writing, but not require production of a complete essay to receive feedback. For example, one practice activity might focus specifically on developing claims and another on refuting counter-claims. Collectively, these solutions should support students’ maintenance of writing performance growth over time and avoid their reaching a ceiling in writing quality prior to demonstrating mastery.

**Limitations of the Present Study**

Findings of the present study should be considered in the context of several limitations. One set of limitations relate to the study design and scope. Due to the number of teachers, schools, and districts involved in the study, and in particular the geographic distance between schools and districts, it was not feasible to observe or record teacher lessons to substantiate survey and interview reports of writing practices. Given the number of student participants a survey rather than interview was used to assess students’ writing knowledge; however, this may have introduced construct-irrelevant variance and conflated students’ writing knowledge with writing ability or motivation.

A second set of limitations were inherent in NC Write. Since NC Write did not include a plagiarism scanner some students produced unoriginal essays and as a result had less writing practice than designed. Lesson activity logs tracked which lessons students
accessed and for how long, but such logs provided no information about student engagement. Additionally, the lessons provided no quantitative measures of performance. Students in the NC + SRSD condition could not be required to use the brainstorming sheet graphic organizer within the software.

A third set of limitations relate to aspects of treatment. Some teachers reported a learning curve associated with NC Write, and in retrospect wished they had spent more time exploring the software prior to implementing the intervention. This would have provided greater familiarity with the lessons, prompts, and other components of NC Write. A few teachers felt the pace of the intervention was a challenge for at least some students; this appeared to be largely related to the amount of writing involved. Some students shared this opinion; one explained, “Sometimes I just felt like I didn’t have enough time to do it, so I just had to rush through it. But it’s partly my fault for stalling, I guess.”

A final set of limitations relate to the timing of the intervention. The intervention has been planned for earlier in the school year but due to the time involved in the district-level IRB review and approval process; meetings with districts, schools, and teachers; the establishment of NC Write accounts for student and teacher users; and teacher training prior to the start of the intervention a single teacher piloted the intervention in the fall and all other teachers implementing in the spring. Since the intervention was primarily conducted in the latter part of the school year teachers were not able to plan around the intervention as well as they would have had it been conducted earlier in the year. For example, one teacher mentioned how some of the NC Write lessons would have held greater utility for her students earlier in the school year. As discussed above, intervention timing may also have resulted in
depressed writing performance towards the end of the intervention and at posttest. The timing also precluded examining writing performance at a maintenance timepoint as intended. Future research should examine whether results are similar if treatment occurs earlier in the school year and as well as examine effects at maintenance.

**Recommendations for Future Research**

In addition to addressing the limitations of the present study, future research should investigate how AWE programs can be improved to best support students’ writing development and teachers’ writing instruction. Future research should continue to examine how AWE programs can be used along with SRSD instruction to combine strategy instruction with the practice opportunities and feedback afforded by AWE programs. Particular consideration should be given to integrating the goal setting aspects of SRSD instruction with automated essay scores. There exists the opportunity to examine how automated essay scores impacts the calibration accuracy of students’ judgments of writing quality and informs goal setting. Accurate judgments are critical for SRL (Labuhn, Zimmerman, & Hasselhorn, 2010).

Future research examining AWE and SRSD interventions should also examine the amount of professional development required for high-fidelity treatment implementation by teachers. Despite the large volume of SRSD writing research produced to date, little research has examined whole-class instruction implemented by teachers. Recent SRSD literature suggests a substantial upfront investment (i.e., 12–14 hours over two days) in practice-based professional development is required (Festas et al., 2015; McKeown et al., 2016); however, there is some evidence that writing growth can be achieved with much less (i.e., 2–3 hours)
teacher training (Ennis et al., 2015). Future research designs should support better understanding of the amount of professional development needed for teachers to successfully implement AWE and SRSD interventions.

The present study did not investigate students’ writing motivation, though interview results suggested NC Write may have positively impacted students’ intrinsic motivation. There is evidence that SRL interventions can positively affect students’ motivation (Dignath et al., 2008), and some evidence that SRSD instruction in particular can positively impact students’ writing motivation (Brunstein & Glaser, 2011; Harris et al., 2015). Immediate feedback is suspected to support motivation but possibly create dependence on feedback information (Shute, 2008). Thus, future research might examine how SRSD instruction and automated feedback uniquely and collectively contribute to students’ writing motivation. An additional line of research might also examine situational interest in the context of AWE programs.

Finally, more research is needed to understand how AWE programs can better support the needs of gifted students. There is mounting evidence from the results of the present study and Grimes and Warschauer (2010) that in general teachers are less positive about AWE programs supporting the needs of gifted writers than other students. Future quantitative research could examine this claim empirically and additional qualitative research could help to better understand teachers’ opinions regarding AWE programs supporting the needs of gifted students.
REFERENCES


APPENDICES
Appendix A
Research Design Overview

Quantitative Data Collection Phase
- Procedures: Collect measures of writing performance (argumentative essay quality, length, number of basic elements), knowledge, and self-efficacy at pre- and posttest from all students (N=20). Collect measures of writing performance during intervention from students in treatment conditions.
- Products: Numeric data

Quantitative Data Analysis Phase
- Procedures: Data screening, ANOVA (SPSS 23), CFA (AMOS 21) multi-level modeling (SAS 9.3)
- Products: Normality of data, descriptive statistics, F-ratio, fit indices, multi-level model coefficients

Qualitative Data Collection
- Procedures: Interview students (n=30) and teachers (n=12) to investigate perceptions of the events, value, and strengths and weaknesses associated with AC Write and the writing intervention.
- Products: Interview recordings, transcripts

Qualitative Data Analysis
- Procedures: Data coding (Nvivo 10), testing and revision of coding scheme
- Products: Coded text, memos, themes

Integration of Quantitative and Qualitative Results
- Procedures: Integration of quantitative findings with qualitative results; examination of complete results
- Products: Discussion, limitations, recommendations for future research
Appendix B
Study Information Email

My name is Corey Palermo and I am a Ph.D. candidate in Educational Psychology at North Carolina State University. I am also a Senior Program Manager at Measurement Incorporated, in Durham, NC. I am working on my dissertation, which examines web-based interventions to support 6-8 grade students’ essay writing. I am looking for 6th, 7th, and 8th grade writing teachers to use a writing program with their students as part of my dissertation study.

Study Overview
The interventions I am examining are part of NC Write, a web-based writing practice program. You may be familiar with this program already; it is currently being used in over 80 schools in 11 districts across North Carolina. NC Write is a formative writing practice program designed to help students develop effective writing skills and maximize teacher instruction. The program provides access to custom writing prompts, electronic graphic organizers, automated scoring, instant feedback, portfolios, and interactive lessons. You can find more information about the program at http://ncwrite.com. My research will examine the impact of the program on students’ writing quality, knowledge, and self-efficacy. I am examining these outcomes specifically related to argumentative essay writing given the emphasis this mode receives in the 6-8 grade standards. My study will have three different conditions. In two conditions, students using the program will be provided with lessons focusing either on writing skills or writing self-regulation. A third condition will serve as comparison condition; students in this group will not use the program and will instead receive regular writing instruction from their classroom teacher. Intact classes will be assigned to conditions.

NC Write
When using NC Write, students will complete a wide variety of interactive lessons. The writing skills lessons address skills needed for effective argumentative essay writing, and focus on the development of ideas, organization, word choice, sentence structure, and conventions. Also included are grammar and passage exercises to support text revisions. The writing self-regulation lessons include strategies for planning and writing argumentative essays, and focus on self-regulatory strategies such as goal setting, self-monitoring, self-instruction, and self-assessment. All lessons are interactive to encourage engagement and retention. With NC Write, all students will have multiple opportunities to practice their argumentative writing by responding to included prompts, will receive instant scores and feedback for each essay submitted, and will have opportunities to revise essays. Collectively, the lessons and activities are designed to provide strategies and practice opportunities necessary to develop proficiency in argumentative writing. An additional purpose of NC Write is to maximize teacher instruction by providing instructional support and removing some of the burden associated with grading and the provision of feedback. NC Write is not meant to replace classroom writing instruction, but augment it. Students will also be asked to complete short pretests and posttests (argumentative essay, writing knowledge, achievement
goal orientation, writing self-efficacy) that align with the program content and argumentative writing.

**Participation**
Participating teachers will be asked to implement the writing intervention with classes in the writing skills and writing self-regulation conditions two days a week for a period of eight weeks. Teachers can continue with their regular writing instruction during the rest of the week. Classes in the comparison condition will receive regular writing instruction for the entirety of the study. Students in all three conditions will be asked to complete the pretests prior to the start of the intervention and posttests at the conclusion of the intervention.

The per-student fee to use NC Write will be waived and student accounts provided free of charge for the duration of the school year for participating teachers.

Having taught in the North Carolina school system myself I realize the value of instructional time. The interventions have been developed to augment classroom writing instruction, and the program designed to maximize teacher instruction. The results of my study will be used to make further improvements to the program with the ultimate goal of better addressing the needs of developing writers.

**Meeting**
I would like to arrange a brief call or meeting with you to discuss NC Write and this study further. Please let me know if you have any questions at this time and if you are willing to schedule a brief conversation in the near future. Thank you for your time.

Corey Palermo
cpalermo@measinc.com
Appendix C
Posttest Directions

Part I: Survey

Before the survey
1. The posttest survey can be accessed using the link:
   http://ncsu.qualtrics.com//SE/?SID=SV_b72zeJR49UzHeBf
2. The posttest survey consist of 3 measures:
   a. A writing self-efficacy measure (Q2)
   b. A measure of writing knowledge (Q3-Q4)
   c. An NC Write student survey (Q5)
3. Inform students that all items should be completed carefully and thoughtfully. Students
   should not skip any items. An error will display if any items, or parts of items, are omitted.

During the survey
4. Ensure students enter their PowerSchool IDs correctly. They will need to re-enter their IDs
   at the end of the survey; an error will display if the IDs do not match.
5. Two items (Q3-Q4) require written responses, and students should be encouraged to write
   down all of their ideas, not just a sentence or two. These items require a minimum
   response. Encourage students to take some additional time to think and then write down
   some more ideas if their response is too brief.

Part II: Essay

6. Students should login to NC Write (ncwrite.com, select “user login”).
7. Tell students to try their hardest and use all of the strategies they have learned when they
   write the essay. Remind students that they will not have an opportunity to revise this essay,
   and should put forth their best effort on the first draft. Tell students that they will still
   receive scores and feedback, but they will not be able to make further changes to their
   essays.
8. Have students select the “practice” tab, select the prompt that has been assigned to
   him/her for the posttest from the left side of the screen (prompt assignments were sent via
   email), then select “begin writing.” Ensure each student has selected the appropriate
   prompt (see screenshot below).
9. When students feel they have completed their essays, have them click “submit,” and confirm they want to submit the essay for scoring.

10. Once the essay has been submitted students may wish to review the feedback, but should NOT revise the essay.

Since this is a posttest, please do not provide students with any assistance during essay writing.
Appendix D
Writing Prompts

Academics for Athletes
Suppose that your school is considering revising the academic requirements for its student athletes. The new policy will require students to maintain a minimum grade of “C” or “Average” in all subjects in order to participate in a sport. Would you be for or against this new policy? Provide facts and details to support your reasons.

Are You Old Enough for a Part-Time Job?
A new law is being considered that would allow students as young as 13 to hold part-time jobs while attending school. Think about the advantages and disadvantages of this law and whether you agree or disagree with it. Write an essay that takes a position for or against this proposed change in the law. Be sure to provide support for your position.

Grading Policy
Your teacher has assigned a group project. Everyone in the group will have a different job, but everyone will get the same grade based on the overall outcome of the project. Write an essay in which you take a position for or against this policy. Use relevant evidence in the form of fact, examples, personal experience or that of someone you know, and your beliefs about fairness.

Let Me Sleep
Write a letter to your superintendent explaining why your school district should make school starting times later for middle and high school students. Use the information in the stimulus material to form and support your argument.

Library Hours
Your town's public library is facing a budget crisis and needs to save money by reducing the amount of time it is open each week. The mayor is considering a proposal to close the library at 5:00 p.m. on weekdays and have it remain closed on weekends. The community has strong opinions about this proposal. You decide to write a letter to the mayor expressing your opinion about the proposal to limit library hours. Write a letter to the mayor expressing your opinion about whether the library should close a 5:00 p.m. on weekdays and remain closed on weekends. Be sure to include reasons, facts, examples, and other evidence to support your position.

Park and Playground Fee
Your state is considering whether to charge a $2 fee for each person to enter and use parks and community playgrounds. The money would be used to maintain and improve buildings fields, courts, and playground equipment. This proposal is controversial, and many citizens have strong opinions about the idea. You decide to write an editorial for the local newspaper expressing your opinion about whether to charge a $2 fee for each person to enter and use parks and community playgrounds. Write an editorial for the local newspaper expressing your opinion about whether to charge a $2 fee for each person to enter and use parks and community playgrounds. Be sure to include reasons, facts, examples, and other
evidence to support your position.

School Concert
A well-known musical group has offered to give a free concert at your school. There has been much debate as to when the concert should be held—during or after school. Write an essay for the school newspaper supporting your position whether the concert should be held during school time or after school. Use reasons, facts, examples, and other evidence to support your position.

Should PE be a Requirement?
Write a letter to the principal arguing for or against the requirement of the physical education class for all students. Make sure your reasons are clear and that you support them with suitable examples and details.

Should the School Year be Extended?
The typical school year is made up of 180 days. Suppose Congress was considering a law to extend the school year to 220 days in order to give students more time to learn and to improve achievement scores. Is this a good or a bad idea? Write a letter to your states' members of Congress, expressing your opinion. Be sure to give strong reasons with plenty of support.

Should We Bring Back Extinct Species?
Recent discoveries with DNA and its regeneration have suggested to scientists that it would be possible to recreate extinct animals from their recovered DNA. Read the articles and view the video about the regeneration of extinct species. You may want to take notes on the video and your reading. Then plan and write an essay in which you argue for or against this idea. Be sure to explain the opposing claim and to cite specific evidence from the three sources to support your claim.

Student Speeches
Suppose the administration of your school is adopting a new policy regarding student speeches presented in school or at school-sponsored events. The policy states that all student speeches must first be reviewed and approved by a faculty committee. Would you be for or against this policy? Provide reasons and detailed support for your reasons.

Talent or Hard Work?
Which is more important: Talent or hard work? Explain your choice and support your reasons with specific details and examples.

1 pretest/posttest prompt
2 collaborative essay prompt
3 source-based prompt
### Appendix E
Sample Scoring Rubric

#### STYLE

Style includes the student’s ownership of the topic and connection with the audience. Style is enhanced through strong word choice and varied sentence constructions that contribute to the writer’s unique voice. Style is addressed in writing standards 1, 2, 3, and 4.

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>The writer personalizes the topic and forms a strong connection with the audience by demonstrating a strong sense of purpose and appropriateness to task and audience.</td>
</tr>
<tr>
<td></td>
<td>- The writer’s voice comes through the composition and strengthens the connection with the audience.</td>
</tr>
<tr>
<td></td>
<td>- Compositional risks are successful and enhance the response.</td>
</tr>
<tr>
<td></td>
<td>- The communication with the audience is strong without seeming forced.</td>
</tr>
<tr>
<td></td>
<td>- The writer establishes and maintains a formal style.</td>
</tr>
<tr>
<td>4</td>
<td>The writer forms a reasonably successful connection with the audience and has awareness of purpose.</td>
</tr>
<tr>
<td></td>
<td>- There is evidence of voice and awareness of the audience.</td>
</tr>
<tr>
<td></td>
<td>- Compositional risks, if evident, are reasonably successful.</td>
</tr>
<tr>
<td></td>
<td>- The communication with the audience is evident and does not seem forced.</td>
</tr>
<tr>
<td></td>
<td>- The writer establishes and maintains a generally formal style.</td>
</tr>
<tr>
<td>3</td>
<td>The writer’s connection with the audience and purpose is uneven.</td>
</tr>
<tr>
<td></td>
<td>- The response may be mundane and lack voice.</td>
</tr>
<tr>
<td></td>
<td>- Compositional risks, if evident, are only partially successful.</td>
</tr>
<tr>
<td></td>
<td>- The writer may not demonstrate awareness of the audience or the writing may seem somewhat artificial.</td>
</tr>
<tr>
<td></td>
<td>- The writing may show an uneven use of formal style.</td>
</tr>
<tr>
<td>2</td>
<td>The writer’s connection with the audience and purpose is weak.</td>
</tr>
<tr>
<td></td>
<td>- The response has a weak awareness of the audience and lacks voice.</td>
</tr>
<tr>
<td></td>
<td>- The response is pedestrian and may be formulaic.</td>
</tr>
<tr>
<td></td>
<td>- Any reaching out to the audience is disruptive to the flow of the response.</td>
</tr>
<tr>
<td></td>
<td>- The response may have insufficient writing to develop a connection with the audience.</td>
</tr>
<tr>
<td></td>
<td>- The writing lacks a formal style.</td>
</tr>
<tr>
<td>1</td>
<td>The writer demonstrates little to no sense of the audience or the purpose of the topic.</td>
</tr>
<tr>
<td></td>
<td>- The response is disconnected from the topic and may have vague or random ideas.</td>
</tr>
<tr>
<td></td>
<td>- The response may be difficult for the audience to understand.</td>
</tr>
<tr>
<td></td>
<td>- The response may have insufficient writing to develop a connection with the audience.</td>
</tr>
</tbody>
</table>
Appendix F
Writing Knowledge Questions

1. When you are asked to write a paper for class or for homework, what kinds of things can you do to help you plan and write your paper?

2. If you were helping a friend write an argumentative essay, what kind of things would you tell him or her to include?
Appendix G
Writing Self-Efficacy Instrument

Rate your degree of confidence by recording a number from 0 to 100 using the scale given below:

<table>
<thead>
<tr>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cannot do at all</td>
<td>Moderately can do</td>
<td>Highly certain can do</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. When writing an essay it is easy for me to get ideas. __________
2. When writing an essay it is hard for me to organize my ideas. (RC) __________
3. When writing an essay it is easy for me to get started. __________
4. When writing an essay I find it easy to make all the changes I need to make. __________
5. When writing an essay it is easy for me to write my ideas into good sentences. __________
6. When writing an essay it is hard for me to keep the paper going. (RC) __________
7. When writing an essay it is hard for me to correct my mistakes. (RC) __________
8. When my class is asked to write an essay, mine is one of the best. __________
Appendix H
Achievement Goals Questionnaire (AGQ)

Scale = 1 (not at all true of me) to 5 (very true of me)

Please use the rating scale to indicate your level of agreement with each statement.

1. It is important for me to do better than other students.
2. It is important for me to do well compared to others in this class.
3. My goal in this class is to get a better grade than most of the other students.
4. I worry that I may not learn all that I possibly could in this class.
5. Sometimes I’m afraid that I may not understand the content of this class as thoroughly as
   I’d like.
6. I am often concerned that I may not learn all that there is to learn in this class.
7. I want to learn as much as possible from this class.
8. It is important for me to understand the content of this course as thoroughly as possible.
9. I desire to completely master the material presented in this class.
10. My goal is to not perform poorly relative to my classmates.
11. I just want to avoid doing poorly in this class compared with others.
12. My goal in this class is to avoid performing poorly compared to the rest of the class.
Appendix I
Survey of Writing Practices

1. What is your gender? Please select the appropriate answer.
   - Male
   - Female

2. What is your ethnicity/race? Please choose one or more.
   - American Indian or Alaska Native
   - Asian
   - Black or African American
   - Hispanic or Latino
   - Native Hawaiian or Other Pacific Islander
   - White

3. What is the highest degree you hold?
   - BA or BS
   - MA or MS
   - Multiple MA or MS
   - Ph.D or Ed.D.
   - Other (please explain)

4. What type(s) of state certification do you currently have? Please choose one or more.
   - Emergency, provisional, or temporary certification
   - Middle School English Language Arts Certification
   - National Board Certification

5. In which of the following grade levels do you currently teach writing? Select all that apply.
   - 6
   - 7
   - 8

6. How many years have you taught writing prior to this year?

7. Please rate your preparation in teaching writing for each of the following. (1 = none, 2 = minimal, 3 = adequate, 4 = extensive)
   - Formal preparation in writing during certification program
• Writing preparation after certification
• Personal writing preparation

8. What percentage of the students you teach writing are best classified as:
   above average writers:
   average writers:
   below average writers:
   in special education:
   minority students:

9. Which best describes your approach to writing instruction?
   • Process approach
   • Skills approach
   • Combination of process and skills approaches
   • Other (please explain)

10. What percentage of the students you teach writing are
    in special education
    minority students

11. Please indicate how often you use each of the following practices during writing
    instruction. (1 = never, several times a year, monthly, several times a month, weekly,
    several times a week, daily, and 7 = several times a day)
    • Verbal praise
    • Direct instruction of skills
    • Teach spelling
    • Teach summarizing
    • Teach strategies for revising
    • Teach strategies for planning
    • Use writing as a learning tool
    • Teach paragraph writing
    • Student self-assessment
    • Sentence combining
    • Imitate models of good writing
    • Teach handwriting
    • Inquiry/research
    • Teach typing
    (0 [never] to 6 [always])
12. Writing adaptations occur only when an activity is done more often with weaker writers than it is with other students in the class. Please indicate how often you use each of the following adaptations for weaker writers. (1 = never, several times a year, monthly, several times a month, weekly, several times a week, daily, and 7 = several times a day)

- Extra encouragement
- Extra time to practice skills/strategies
- Extra time to do assignments
- Extra review of skills/strategies
- Extra sentence writing instruction
- Extra grammar instruction
- Extra conferencing
- Extra mini-lessons
- Extra revising instruction
- Reteach skills/strategies
- Extra instruction on text structure
- Extra capitalization instruction
- Extra spelling instruction
- Extra opportunities to write with peers
- Extra planning instruction
- Alternative writing assignments
- Select own topic
- Extra opportunities to use work processing
- Extra handwriting instruction
- Extra writing instruction via technology
Appendix J
NC Write Surveys

NC Write teacher survey

1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree

1. NC Write gives fair and accurate scores.

Comments:

2. Please indicate your agreement or disagreement: NC Write makes writing instruction easier.
   saves me time.
   makes teaching more enjoyable.
   lets me focus on higher concerns of writing instead of mechanics.

Comments:

3. Compared to writing with a word processor program such as Microsoft® Word, when using NC Write my students are more motivated to write.
   revise more.
   write more mechanically and superficially.

Comments:

1 = Very Negative; 2 = Slightly Negative, 3 = Neutral, 4 = Slightly Positive, 5 = Very Positive

4. For each category of students, please indicate how much NC Write assists or impedes their writing development.
   English Language Learners
   Special Education
   Gifted
   At-risk
   General students, no special needs

Comments:
NC Write student survey

1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree

Please use the rating scale to indicate your level of agreement with each statement.

1. I find NC Write easy to use.
2. I sometimes have trouble knowing how to use NC Write.
3. I like writing with NC Write.
4. I revise my writing more when I use NC Write.
5. Writing with NC Write has increased my confidence in my writing.
6. NC Write provides good suggestions for improving my writing.
7. The essay scores NC Write gives are fair.
8. NC Write helps improve my writing.
Appendix K
Interview Guides
Teacher Interview Protocol: NC Write + SRSD instruction

1. Has the NC Write program helped your students become better writers? How?
2. Did the feedback provided by the NC Write program help your students improve their writing?
3. What would you change about the NC Write program?
4. Would you recommend the STOP and DARE strategies to other teachers?
5. Do you plan to teach or use these strategies in the future? Is there anything you would change?
6. What did you like about the writing or self-regulation (goal setting, self-instructions, self-monitoring, self-reinforcement) strategies?
7. What did you dislike about the writing or self-regulation strategies?
8. Do you feel the writing intervention was appropriate for a variety of students?
9. Would you recommend the writing intervention to other teachers?
10. Are there any other comments you would like to make about NC Write or the writing intervention?
Teacher Interview Protocol: NC Write + Traditional writing instruction

1. Describe your approach to teaching argumentative writing.

2. Has the NC Write program helped your students become better writers? How?

3. Did the feedback provided by the NC Write program help your students improve their writing?

4. What would you change about the NC Write program?

5. Do you feel the writing intervention was appropriate for a variety of students?

6. Would you recommend the writing intervention to other teachers?

7. Are there any other comments you would like to make about NC Write or the writing intervention?
Student Interview Protocol: NC Write + SRSD instruction

1. Has using the NC Write program helped you become a better writer? How?

2. What did you learn from using the NC Write program?

3. How do you think the NC Write program could help other students?

4. What would you change about the NC Write program?

5. Did the feedback provided by the NC Write program help you improve your writing? Why or why not?

6. Do you think STOP and DARE should be taught to other students? Why or why not?

7. Did you like the STOP strategy? Why or why not?

8. Did you like the DARE strategy? Why or why not?

9. Did you like setting goals? Why or why not?

10. Did you like using self-statements? Why or why not?
Student Interview Protocol: NC Write + Traditional writing instruction

1. Has using the NC Write program helped you become a better writer? How?

2. What did you learn from using the NC Write program?

3. How do you think the NC Write program could help other students?

4. What would you change about the NC Write program?

5. Did the feedback provided by the NC Write program help you improve your writing?
<table>
<thead>
<tr>
<th>Concept</th>
<th>Dimensions/Properties</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Benefits of NC Write</td>
<td>1.1 Efficiency</td>
<td>Perceptions of benefits of NC Write for learning and teaching</td>
</tr>
<tr>
<td></td>
<td>1.2 Evidence of growth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.3 Provides a structure for writing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4 Relevance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5 Supports differentiation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.6 Supports intrinsic motivation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.7 Supports writing instruction</td>
<td></td>
</tr>
<tr>
<td>2. Change about NC Write</td>
<td>2.1 Opinions</td>
<td>Aspects of NC Write users felt should be improved or added</td>
</tr>
<tr>
<td>3. Writing quality feedback</td>
<td>3.1 How feedback supported improvement</td>
<td>Perceptions, implementation, and use of automated feedback</td>
</tr>
<tr>
<td></td>
<td>3.2 How students used feedback</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.3 Teacher assistance with</td>
<td></td>
</tr>
<tr>
<td></td>
<td>interpretation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.4 Criticisms of feedback</td>
<td></td>
</tr>
<tr>
<td>4. NC Write Lessons</td>
<td>4.1 How supported improvement</td>
<td>Perceptions of interactive lessons</td>
</tr>
<tr>
<td></td>
<td>4.2 Criticisms</td>
<td></td>
</tr>
<tr>
<td>5. Prompts</td>
<td>5.1 Perceptions</td>
<td>Perceptions of intervention prompts</td>
</tr>
<tr>
<td>6. Graphic organizers</td>
<td>6.1 How supported improvement</td>
<td>Perceptions of graphic organizers in NC Write</td>
</tr>
<tr>
<td></td>
<td>6.2 Criticisms</td>
<td></td>
</tr>
<tr>
<td>7. STOP and DARE strategies</td>
<td>7.1 General perceptions</td>
<td>Perceptions and implementation of SRSD strategies</td>
</tr>
<tr>
<td></td>
<td>7.2 How supported writing quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.3 Teaching of strategies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.4 Criticisms</td>
<td></td>
</tr>
<tr>
<td>8. Self-statements</td>
<td>8.1 How students used</td>
<td>Perceptions and use of self-statements</td>
</tr>
<tr>
<td></td>
<td>8.2 General perceptions</td>
<td></td>
</tr>
<tr>
<td>9. Goal setting</td>
<td>9.1 How supported writing quality</td>
<td>Perceptions and use of goal setting</td>
</tr>
<tr>
<td></td>
<td>9.2 How students used goals</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.3 Criticisms</td>
<td></td>
</tr>
<tr>
<td>10. Explanation for</td>
<td>10.1 Ceiling on improvement</td>
<td>Factors explaining growth saturation</td>
</tr>
<tr>
<td>quadratic growth</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix M
SRSD Materials

STOP and DARE Chart

**STOP**

S - Suspend judgment  
T - Take a side  
O - Organize ideas  
P - Plan more as you write  

Did I list ideas for both sides?  

Can I think of anything else? Try to write more.  

Another point I haven’t considered is…. Think of possible arguments.  

Put a star next to ideas you want to use. Choose at least five ideas that you will use.  

Put an X next to counterclaims you want to dispute.  

Number your ideas in the order you will use them.

**DARE**

D - Develop your claim  
A - Add supporting reasons and evidence  
R - Reject counterclaims  
E - End with a conclusion
STOP and DARE Checklist

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspend Judgment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Did I list ideas for both sides?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Can I think of anything else?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Try to write more.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Another point I haven’t considered is...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Think of possible arguments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Take a Side</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Place a + at the top of one column to show the side you will take.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organize Ideas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Put a star next to ideas you want to use.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Choose at least five ideas that you will use.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plan More as You Write</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Use DARE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Record the number of sentences for each part:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Develop Your Claim</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Add Supporting Reasons and Evidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reject Counterclaims</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>End with a Conclusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Brainstorming to Choose a Claim

Use this organizer to help you decide on your claim.

**Topic or issue in the form of a question:**

<table>
<thead>
<tr>
<th>For</th>
<th>Against</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1.</td>
</tr>
<tr>
<td>2.</td>
<td>2.</td>
</tr>
<tr>
<td>3.</td>
<td>3.</td>
</tr>
<tr>
<td>4.</td>
<td>4.</td>
</tr>
<tr>
<td>5.</td>
<td>5.</td>
</tr>
<tr>
<td>6.</td>
<td>6.</td>
</tr>
<tr>
<td>7.</td>
<td>7.</td>
</tr>
<tr>
<td>8.</td>
<td>8.</td>
</tr>
</tbody>
</table>
Self-Statements

To think of good ideas:

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________

While I work:

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________

To check my work:

________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
________________________________________________________________
### Transition Words

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adding Information</td>
<td>And</td>
</tr>
<tr>
<td></td>
<td>Another</td>
</tr>
<tr>
<td></td>
<td>In addition</td>
</tr>
<tr>
<td></td>
<td>Also</td>
</tr>
<tr>
<td></td>
<td>Furthermore</td>
</tr>
<tr>
<td></td>
<td>Likewise</td>
</tr>
<tr>
<td></td>
<td>In fact</td>
</tr>
<tr>
<td></td>
<td>Besides</td>
</tr>
<tr>
<td></td>
<td>First, Second, Third, Fourth...</td>
</tr>
<tr>
<td></td>
<td>The following</td>
</tr>
<tr>
<td></td>
<td>Finally</td>
</tr>
<tr>
<td>Ordering Ideas</td>
<td>For example</td>
</tr>
<tr>
<td></td>
<td>For instance</td>
</tr>
<tr>
<td>Providing Examples</td>
<td>Because of</td>
</tr>
<tr>
<td></td>
<td>Due to</td>
</tr>
<tr>
<td></td>
<td>Since</td>
</tr>
<tr>
<td>Providing Reasons</td>
<td>As a result</td>
</tr>
<tr>
<td></td>
<td>Therefore</td>
</tr>
<tr>
<td>Cause and Effect</td>
<td>But</td>
</tr>
<tr>
<td></td>
<td>However</td>
</tr>
<tr>
<td></td>
<td>Despite</td>
</tr>
<tr>
<td></td>
<td>Although</td>
</tr>
<tr>
<td>Contrasting Ideas</td>
<td>To summarize</td>
</tr>
<tr>
<td></td>
<td>In conclusion</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Number of ideas I will use</td>
<td></td>
</tr>
<tr>
<td>Number of essay parts I will include</td>
<td></td>
</tr>
<tr>
<td>Record the number of sentences for each part:</td>
<td></td>
</tr>
<tr>
<td>Develop Your Claim: Goal</td>
<td></td>
</tr>
<tr>
<td>Add Supporting Reasons and Evidence: Goal</td>
<td></td>
</tr>
<tr>
<td>Reject Counterclaims: Goal</td>
<td></td>
</tr>
<tr>
<td>End with a Conclusion: Goal</td>
<td></td>
</tr>
<tr>
<td>Total Score (out of 30): Goal</td>
<td></td>
</tr>
</tbody>
</table>
Should Students Listen to Music While They Study?

Everywhere you go you see students wearing headphones or earbuds, listening to music. Some students even wear them in the library as they study or complete homework. But, is listening to music while studying a good idea? To achieve the greatest learning benefits, students should not listen to music while studying or completing schoolwork.

Listening to music while studying is distracting. Students often study by reading and writing notes, and these activities take a lot of mental effort. Researchers have shown that humans’ brains can’t focus on two things at the same time. Students trying to study may end up singing to their favorite songs instead of concentrating on what they are supposed to be studying. Most likely, students will shift their concentration back and forth between what they are studying and what they are listening to. As a result it will take more time to study effectively than studying without the distraction of music.

In addition, music can impact how deeply students learn while studying. Some studies have shown that students who multitask while learning remember less than students who don’t multitask. For example, researchers have shown that students who are heavy multitaskers are more likely to skim what they are reading instead of analyzing the text carefully. Furthermore, these students have been shown to earn worse grades than their peers. Therefore, students who do other things while studying, like listening to music, don’t learn very well due to the distractions.

Some students feel that listening to music can be helpful when trying to reduce other distractions. For example, a busy library can be rather noisy and students may listen to headphones to block out that noise. However, a better option would be to find a quiet location to study. Some libraries have quiet sections which are more suitable for studying. Also, a pair of earplugs would be a better option than music for students who must study in busy or noisy locations.

In conclusion, there are many distractions available to students when studying, including music. To get the most of out their studying and schoolwork, students should not distract themselves with music as it limits how effectively they learn. Instead of listening to music while studying, why not reward yourself with music when you have finished studying and can really enjoy it?
Should Schools Teach about Recycling?

Some schools have recycling programs where students will find different colored bins placed around the building for plastic, glass, and paper products. Yet some schools choose not to use instruction time to inform students about recycling. Are bins placed around the school building enough for students to understand the importance of recycling? No! Schools should teach students to reduce, reuse, and recycle.

As a nation, the United States produces a lot of trash. This trash ends up in landfills that are quickly becoming full. For example, once plastic ends up in a landfill it may take up to 1,000 years to decompose. Plastic sitting in a landfill may pollute the soil and water, causing damage to the ecosystem. This is not a good solution to getting rid of products people no longer want!

Another reason that recycling should be taught in school is because since school is required, it is the ideal place to introduce students to recycling. Teachers could not only explain about the toxic chemicals, but also give students a way to get involved with making the planet a better place to be. In school students would not only learn about recycling and practice recycling, they would also take what they’ve learned home and share it with their families. They would learn how to become responsible for our planet.

An additional way to recycle is to reuse items. A plastic grocery bag can become tomorrow’s lunch bag. Plastic water bottles can be refilled. If people reuse products that would otherwise be thrown away, companies won’t have to make as many new products. Also, some new products can be made out of recycled materials so there is less waste. For example, bottles can be made from recycled glass.

Some people have pointed out that a lot of the plastic that could be recycled doesn’t actually end up being recycled. In 2011, the Environmental Protection Agency reported that 32 million tons of plastic waste was made but only 8% of plastic was recycled. The rest of the plastic that was used ended up in landfills. While that report is discouraging, it is not a reason to give up. We only have one planet and it is worth taking care of.

Think of all the plastic that comes with the products you buy: bags, bottles, jugs, cups, containers, and other types of packaging. If placed in the trash these plastics end up in landfills. While decomposing over centuries, these plastics may pollute the land and water. There’s a Native American Arapaho proverb that says, “Take only what you need and leave the land as you found it.” One way to do this is by recycling.

Should schools teach about recycling? Absolutely. We need to take care of our planet since our planet provides for us. We have a responsibility to our planet and to future generations of humans. School students are in the best position to make a change. Our planet deserves better and so do we.
Should the School Day Start Later?

Most middle and high school students begin their school day before 8:30 AM. There is a lot of debate as to whether it is best for middle and high school students to start learning so early in the morning. Scientific evidence now available confirms that the school day should start after 8:30 AM for middle and high school students.

First, middle and high school students are experiencing a lot of biological changes. Therefore, they require more sleep than younger children. According to the National Sleep Foundation, teenagers need an average of nine and one quarter hours of sleep a night. Since they need to get up early, maybe the problem is that middle and high school students stay up too late. Unfortunately, homework, family responsibilities, and extra-curricular activities like sports and clubs take up a lot of after-school time, and prevent students from going to bed early.

Second, tired students are not effective learners. When the school day starts early in the morning students may be struggling to keep their eyes open, not alert and attentive. Drowsy students are not likely to pay attention in class or participate meaningfully in activities. This limits how much and how well students learn.

Third, The National Sleep Foundation showed that lack of sleep in teens was associated with depression, poor impulse control, and poor decision making. A later start to the school day would give middle and high school students a better opportunity to get the sleep they need to be successful.

With all of this evidence, why does the school day still start so early for many middle and high school students? Ted Velkoff of USA Today offers some explanations in the article “Later Start Times Cost: Opposing View.” One reason is the issue of transportation. School systems develop bus schedules so that elementary, middle, and high school students can be bussed to and from school, usually on the same busses. If middle and high schools started at the same time as elementary schools more busses would be needed to transport all students to school. A second reason is the issue of scheduling after-school sports. Currently, some outdoor sporting events go right up until dusk. These outdoor sports events can’t start any later or student athletes would be playing them in the dark!

Despite these scheduling and cost challenges, school systems should look carefully at the evidence that has been provided by the scientific community. A solution may not be cheap, but aren’t middle and high school students worth it? And shouldn’t school systems give students every opportunity to be successful? This means starting the school day after 8:30 AM for middle and high school students.
Should Kids Participate in Community Service Programs?

Whether you live in the country or the city, in the mountains, close to the beach, or somewhere in between, you have probably seen trash littering local streets, a stray animal that needs a home, or a person in need of assistance. What do these things all have in common? There are community service programs that provide assistance in all of these situations. What’s more, there are a range of benefits associated with volunteering. Kids should participate in community service programs that fit their interests and availability.

Lots of kids are looking for ways to make an impact on the world around them. Helping out in the community is a great way to start. Some schools have afterschool or summer programs that match kids with volunteer and community service opportunities. Some of these programs operate year-round, while others are active at certain times of the year, like holidays. There are many options for kids to make a difference in their own community.

The website www.unitedway.org offers lots of information on different ways for kids to participate in community service programs. For example, kids can volunteer at summer camps, local parks, nursing homes, or hospitals. The benefits of these experiences are that kids can expand their social circles, learn empathy and compassion for others, set life-long patterns of advancing the common good, and increase their chances of success in life. Participating in a community service program lets kids learn about commitment, responsibility, and work ethic, and offers results kids can see and feel proud of. That sounds better than sitting on the couch watching TV or playing video games, doesn’t it?

Some kids might feel they are too busy to participate in community service projects. Middle school students have lots of responsibilities, and there are only so many hours in the day. Also, shouldn’t schoolwork be kids’ focus, and not other activities like volunteering? You might be surprised to learn that many colleges and employers look for students who have been involved in community service programs. Students may even be eligible for a college scholarships based on their community service experience.

Another reason that kids should participate in community service programs is that volunteering can bring families closer together, according to www.kidshealth.org. Participating in a community service program could become a family tradition, and give kids the opportunity to spend more time with their families while giving back to the community. With their parents or other family members, kids might work in a soup kitchen, volunteer at an animal shelter, or pick up trash in a local park.

In conclusion, all kids should participate in community service programs. Volunteering offers benefits for kids while helping the community at the same time. And that’s a win-win situation.
Should Textbooks be Replaced with Notebook Computers?

More and more notebook computers are being used in schools across the country. Some schools are providing students with notebook computers instead of paper textbooks. Is this a good idea? Some kids think so, but notebook computers can also raise a lot of problems. I think the problems are worth it, though. Notebook computers are a great idea and should be used in every classroom.

Textbooks are really heavy. Kids shouldn’t be expected to drag textbooks back and forth to school every day on the bus. They take up a lot of room and hurt your back.

It is hard to find what you are looking for in a textbook. With a notebook computer you can find what you are looking for instantly. They are really awesome. You can also take notes on them in class and have everything in one place.

Notebook computers allow you to do research right there in the classroom. No more walking single file down the hallway to go to the media center. You have everything you need right there at your fingertips.

Notebook computers are great and should be used in every classroom.
<table>
<thead>
<tr>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>STEP 1: Suspend judgment</strong>&lt;br&gt;Did I list ideas for both sides?</td>
<td><strong>STEP 1: Suspend judgment</strong>&lt;br&gt;Can I think of anything else? Try to write more.</td>
<td><strong>STEP 1: Suspend judgment</strong>&lt;br&gt;Another point I haven’t considered is.... Think of possible arguments.</td>
<td><strong>STEP 2: Take a side</strong>&lt;br&gt;Place a + at the top of one column to show the side you will take.</td>
</tr>
<tr>
<td>#5</td>
<td>#6</td>
<td>#7</td>
<td>#8</td>
</tr>
<tr>
<td><strong>STEP 3: Organize ideas</strong>&lt;br&gt;Put a star next to ideas you want to use. Choose at least five ideas that you will use.</td>
<td><strong>STEP 3: Organize ideas</strong>&lt;br&gt;Did I star ideas on both sides? Choose at least two counterclaims that you can dispute.</td>
<td><strong>STEP 3: Organize ideas</strong>&lt;br&gt;Number your ideas in the order you will use them.</td>
<td><strong>STEP 4: Plan more as you write.</strong>&lt;br&gt;Develop your claim.&lt;br&gt;Add supporting reasons and evidence.&lt;br&gt;Reject counterclaims.&lt;br&gt;End with a conclusion.</td>
</tr>
</tbody>
</table>