Strategy Instruction with Self-Regulation in College Developmental Writing Courses:

Results from a Randomized Experiment

Charles A. MacArthur[1], Zoi A. Traga Philippakos[2], Henry May[1], and Jill Compello[1]

[1]University of Delaware

[2]University of Tennessee

Author Note

Correspondence concerning this article should be addressed to Charles A. MacArthur, School of Education, University of Delaware, Newark, DE 19716. macarthu@udel.edu

Abstract

The paper presents the results of a randomized experimental study of a writing curriculum for college developmental writing courses based on strategy instruction with self-regulation integrated with practices common in college composition. Students in a full semester course learned strategies for planning and revising based on rhetorical analysis and genres. In addition, they learned metacognitive, self-regulation strategies for goal setting, task management, self-evaluation, and reflection. A prior quasi-experiment found positive effects of the curriculum on writing quality, self-efficacy, and mastery motivation. The current study included 19 instructors and 207 students across two colleges. Using hierarchical linear modeling (HLM) with students nested within instructors and with condition and college as factors and pretest scores as covariates, analyses found positive effects of the treatment for quality of argumentative writing ($ES = 1.18$), quality of writing on an independent writing assessment ($ES = .67$), and several motivation outcomes, including self-efficacy (for tasks and processes ($ES = .50$), for grammar ($ES = .36$), and for self-regulation ($ES = .40$)), affect ($ES = .32$), and beliefs about the importance of content ($ES = .29$). No significant effects were found for grammar/conventions or reading comprehension. Teachers in the treatment condition commented positively on the approach and noted improvements in student writing and motivation. Students also shared positive experiences and noted improvement in their writing.

Keywords: basic writing, writing instruction, strategy instruction, metacognition, self-regulation

Educational Impact and Implications Statement

This study demonstrates strong positive effects of strategy instruction with self-regulation in college developmental writing courses. Students learned strategies for planning and revising along with metacognitive strategies for managing their learning. Results revealed substantial improvement in students' wring quality and motivation. The findings are significant for administrators and instructors looking for evidence-based practices and for policy makers as they consider reforms in community college developmental programs.

Strategy Instruction with Self-Regulation in College Developmental Writing Courses:

Results from a Randomized Experiment

In the United States, many students newly enrolled in postsecondary programs are evaluated as under-prepared for college and required to take non-credit remedial courses in writing, reading, and/or math, termed developmental courses. Developmental courses are offered in 99% of public community colleges and 74% of public four-year colleges (Synder et al., 2019). A transcript study from the National Center for Educational Statistics (Chen et al., 2016) of students beginning college in 2003 found that 59% of community college (CC) students and 33% of 4-year college students took developmental math while 28% of CC students and 11% of 4-year students took developmental writing or reading. Assignment to developmental education courses was more common for Black and Hispanic students. Colleges often offer multiple levels of developmental writing courses, so that students placed in lower levels need to take two or even three developmental courses before enrolling in first-year composition (FYC). Overall, students who are placed in developmental writing courses are less likely to pass FYC than students placed directly into FYC (Chen et al., 2016; Attewell et al., 2006). For example, one study of students referred to developmental English/reading courses found that only 37% of referred students successfully completed a related credit course (Bailey et al., 2010). Differences by race/ethnicity and gender were also found; Black and Hispanic students, compared to White students, were less likely to successfully complete developmental courses and pass FYC, and the same was true for male compared to female students (Bailey et al., 2010; Chen et al., 2016). A study at a minority-serving community college (Nastal-Dema, 2019) found that only 12% of students who took the lowest level developmental English course passed FYC.

A number of studies have attempted to improve outcomes through structural changes in the organization of developmental courses. One type of structural change is combining multiple developmental courses into a single course to help students complete requirements sooner. Studies of combining two developmental writing courses (lower and higher levels) into a single course (Hodara & Jaggars, 2014) and studies of integrating developmental reading and writing courses (Edgecombe et al., 2014) have found positive effects on success in later courses. Another option involving structural change is co-requisite courses, which offer students the opportunity to take FYC simultaneously with a developmental course designed to support success in FYC. Research (Cho et al., 2012) on the original co-requisite model, the Accelerated Learning Program (Adams et al., 2009), has found that students are more likely to take and pass FYC in a co-requisite program than traditional developmental education.

Despite the prevalence of developmental writing courses in community colleges and the interest in structural changes in those courses, little research has focused on instructional methods within developmental courses. A review of research on developmental literacy classes (Perin, 2013) found only 13 experimental or quasi-experimental studies, none of which focused on approaches to writing instruction. A more recent review (Perin & Holschuh, 2020) of research on literacy instruction in developmental education found three studies of reading instruction, but the only experimental study of writing instruction was part of our research program, a quasi-experimental study of the Supporting Strategic Writers (SSW) instructional approach (MacArthur, Philippakos, & Ianetta, 2015). The current study extends research on the SSW approach with a rigorous randomized control trial (RCT); thus, it addresses the need for research on instructional methods in developmental writing courses.

Theoretical Framework

Writing is a complex cognitive and social process that draws on writers' content knowledge, discourse and rhetorical knowledge, cognitive and metacognitive strategies, motivation, and basic writing skills (Bazerman, 2016; Graham, 2018; Hayes, 1996; Bereiter & Scardamalia, 1987; Zimmerman & Risemberg, 1997). Sociocultural models of writing (Bazerman, 2016) emphasize that the knowledge and motivation required for writing are based on social purposes and are learned through social interaction in and out of school. Thus, writers need knowledge of the social and rhetorical purposes of writing and the genres commonly used for those purposes; this knowledge guides writers' cognitive processes.

Cognitive models of writing (Bereiter & Scardamalia, 1987; Hayes, 1996; MacArthur & Graham, 2016) emphasize that writing is a goal-directed process intended to meet social purposes. When planning, proficient writers set goals based on rhetorical purposes and select genres appropriate to those purposes; they seek information through reading and note-taking, organize their texts, and apply these processes recursively. They also engage in substantial revision throughout the writing process, guided by their rhetorical goals and knowledge of evaluation criteria appropriate for varied writing tasks (Hayes, 2004; MacArthur, 2016). In contrast, beginning or less proficient writers do far less planning and revising (McCutchen, 2006). Because of the complexity of writing processes, proficient writers use self-regulation strategies to manage the writing process and their personal motivation and behavior (Harris et al., 2018; Zimmerman & Risemberg, 1997). For example, writers engage in goal setting, self-monitoring, managing time and workplace, self-reinforcing, and self-evaluation. Self-regulation is especially important for motivationally challenging tasks like writing.

Cognitive models of writing also include translation and transcription processes. Translation (Fayol, 2016) refers to generation of language to represent ideas, and transcription involves skills of spelling, handwriting, and typing needed get language into print. These basic writing skills are important because problems can interfere with fluent writing and with attention to higher level writing processes (Abbott et al., 2010). In addition, errors in grammar and spelling can affect communication in real life situations (Graham et al., 2011). Research on grammar instruction has shown that traditional grammar instruction has little to no impact on writing quality for native speakers (Hillocks, 1984; Graham & Perin, 2007; Andrews et al., 2006), though the results are more positive for students learning a second language (Ferris, 2012). Some research has shown positive effects on writing quality of sentence-combining approaches to improving sentence structure (Andrews et al., 2006) and of grammar instruction embedded with overall writing instruction (Jones et al., 2013).

The motivational demands of writing are substantial, and motivation is included in both sociocultural (Bazerman, 2016) and cognitive models (Hayes, 1996; 2006; Zimmerman & Risemberg, 1997). Motivation to engage in writing generally declines across the school years (Hidi & Boscolo, 2006), and it may be especially problematic for students encountering the demands of college writing for the first time. A qualitative study of first-year composition students in a community college (Cox, 2009) found that 80% of students mentioned fears of failure, and that those fears often led to maladaptive behaviors like withdrawing from classes or not submitting assignments for evaluation. Motivation for writing is a complex concept that has been studied from multiple theoretical frameworks (MacArthur & Graham, 2016). The largest body of studies on writing motivation has focused on self-efficacy, or confidence in one's ability to complete specific tasks (Bruning & Kauffman, 2016; Pajares & Valiante, 2006). In writing,

self-efficacy has been shown generally to predict writing performance. Research has consistently found that mastery experience is the strongest source of self-efficacy beliefs, and one study that addressed writing confirmed this finding (Pajares et al., 2007). Another theoretical framework that has been applied to writing motivation is the trichotomous model of achievement goals, including mastery, performance-approach, and performance-avoidance goals (Pintrich, 2000). Individuals with a mastery goal orientation seek to develop knowledge and competence. Those with a performance-approach goal orientation aim to perform well compared to others. Performance-avoidance goals refer to efforts to avoid unfavorable judgments, efforts which can limit engagement in learning as students seek to hide their weaknesses. Research in writing has found positive correlations of writing quality with mastery goals and negative correlations with performance-avoidance (Kauffman et al., 2010; MacArthur, Philippakos, & Ianetta, 2015; Traga Philippakos et al., 2015).

Two other motivational constructs that have been studied less often are affect and beliefs about what contributes to good writing. Affect, defined simply as liking writing, is potentially important since people are more likely to engage with things they like; research has found positive correlations of affect with other motivational variables, including self-efficacy and mastery goals (Bruning et al., 2013; MacArthur, Philippakos, & Graham, 2016). White and Bruning (2005) argued that beliefs about writing as a form of meaningful communication affect motivation; they found that higher achieving writers had stronger beliefs in writing as a way to develop understanding and lower beliefs in writing as a means of transmitting the knowledge of others. Research with basic writers (Shaughnessy, 1977) has found that beliefs that overemphasize errors create anxiety and restrict writing growth. Recent research (MacArthur, Philippakos, & Graham, 2016) has found that beliefs in the importance of content are positively

related to writing quality while beliefs focused on the importance of conventions are negatively related to quality.

Writing development is also closely related to reading development. Reading and writing draw on common knowledge and cognitive processes at all levels (Fitzgerald & Shanahan, 2000; Shanahan, 2016). Reading comprehension and composing both draw on knowledge about content, discourse, and strategies. Meta-analytic reviews of research have found substantial effects of teaching writing on reading (Graham & Hebert, 2011) and conversely substantial effects of reading instruction on writing (Graham et al., 2018).

Strategy Instruction with Self-Regulation

The central idea of strategy instruction is that it is possible to teach students to use strategies based on the cognitive and metacognitive strategies used by proficient readers or writers (or experts in any area) (MacArthur, 2011). Substantial bodies of research with elementary and secondary students have demonstrated positive effects of strategy instruction on reading comprehension (Dole et al., 2009) and writing (Graham, McKeown et al., 2012; Graham & Perin, 2007). One of the most effective models of strategy instruction is Self-Regulated Strategy Development (SRSD; Harris & Graham, 2009), which integrates domain-specific strategies (e.g., writing) with self-regulation strategies. A review of multiple meta-analytic reviews of writing instruction (Graham et al., 2016) showed that adding self-regulation strategies to writing strategies substantially increases the effects on writing outcomes.

Strategy instruction draws on multiple areas of research to address three instructional design questions: what strategies to teach, how to support independence and self-regulation, and what pedagogical methods to use (Harris & Graham, 2009; MacArthur, 2011).

First, decisions about the design of writing strategies draw on research on the knowledge and cognitive processes of proficient writers (Hayes, 1996) including knowledge about social contexts and genres. Proficient writers engage in rhetorical analysis of audience and purpose and apply substantial discourse knowledge about the genres appropriate to varied purposes and contexts. Proficient writers also have a repertoire of strategies for planning, drafting, evaluating, and revising. Most writing strategy research uses strategies that teach students how to use discourse knowledge about particular genres strategically to plan or revise (Englert et al., 1991; Graham & Harris, 2012).

Second, theories of self-regulation inform instructional decisions about what is needed to help learners use strategies independently and to adapt them thoughtfully to take control of their own learning (Harris & Graham, 2009; Zimmerman & Risemberg, 1997). Research has found support for instruction in specific self-regulation strategies, including self-monitoring, self-evaluation, self-instructions, goal setting, self-reinforcement, and management of time and environment (Santangelo et al., 2016; Schunk & Zimmerman, 2007). Goal setting, in particular, was found to have positive effects on writing quality in a meta-analysis of writing instruction (Graham et al., 2016). Explicit attention to the development of self-regulation strategies is a key component of the SRSD model (Harris & Graham, 2009).

Third, the pedagogical methods used to teach strategies are supported by a large body of research (Duke & Pearson, 2009; Harris & Graham, 2009; MacArthur, 2011; Pearson & Gallagher, 1983). Teachers establish a meaningful context for learning, provide explicit explanations of the strategies and why and when to use them, use think-aloud modeling to make the process visible, guide student practice, and gradually release responsibility. The SRSD model (Harris & Graham, 2009) provides an especially clear set of steps to guide instruction that

include development of background knowledge, discussion of the strategy, think-aloud modeling, memorization of the strategy, supported practice, and independent practice.

Supporting Strategic Writers

Supporting Strategic Writers (SSW) is based on strategy instruction with self-regulation. The learning goals are common to college composition -- that students will develop rhetorical knowledge of audience, purpose, and genres; strategies for planning and revising; and motivational beliefs that support continued growth in writing (Rose, 1989; CWPA, 2014). In this section, we summarize the writing strategies, self-regulation strategies, and pedagogical methods used in SSW.

The writing strategies are based on rhetorical analysis and genres, with a focus on genres often taught in college, including procedural and causal explanation, personal narrative, comparison, and most important, argumentation (ACT, 2012). The writing strategies draw on the strategies for expository writing developed as part of the Cognitive Strategies Instruction in Writing (CSIW) program of Englert and colleagues (1991). Like the CSIW strategies, the SSW strategies use genre elements to integrate planning and revising. The planning strategies ask students to analyze the writing task rhetorically for audience and purpose, brainstorm ideas, choose an appropriate genre (e.g., comparison, argument), and use a graphic organizer with the genre elements to generate and organize content. The revising strategy asks students to apply an evaluation rubric based on the same genre elements.

Students use the evaluation rubric for self-evaluation and peer review. Two meta-analytic reviews (Graham et al., 2015; Hillocks, 1986) found positive effects on writing quality from instruction in evaluation criteria with practice applying the criteria to make revisions. Research reviews have also found positive effects of peer review with college students and adolescents

(Graham & Perin, 2007; MacArthur, 2016; Topping, 1998). Peer review is a reciprocal process in which students both give and receive feedback. Giving feedback itself may be helpful because it requires students to apply evaluation criteria and give suggestions for revision. Several studies have found positive effects of instruction in evaluation rubrics followed by opportunities to give feedback to peers (Cho & MacArthur, 2011; Lundstrom & Baker, 2009; Philippakos & MacArthur, 2016).

Instruction in grammar and conventions is embedded in the process of revising and editing. The evaluation criteria include an emphasis on writing clear sentences, especially key sentences such as the thesis and topic sentences, which might improve sentence structure. In addition, editing lessons for each assignment include guided practice in fixing typical errors in the context of complete papers, followed by peer editing and individual teacher feedback.

The self-regulation strategies in SSW are based on SRSD with adaptations for the college setting. Drawing on the theme of 'academic success' in community colleges (Bickerstaff et al. 2017), our self-regulation strategies are presented as a set of *Strategies for Academic Success* (SAS): goal setting, task management, progress monitoring, and reflection. For each writing assignment, students set goals based on past performance, plan when and where they will work, monitor their use of the strategies, and reflect on their performance before setting new goals. As in SRSD, instructors discuss the importance of taking control of the writing process and include self-regulation comments in their think-aloud modeling; however, students are not taught to use self-statements. Instead, students reflect in journals about their goals, strategies, and writing progress, and then participate in class discussions about their reflections. The class discussions provide group support as students are encouraged to understand that they can take control of the writing process and begin to see themselves as writers (Oyserman, 2007).

The pedagogical methods used in SSW are based on research on strategy instruction. Like SRSD, SSW uses a clear sequence of steps that functions as a strategy for teaching strategies (Harris & Graham, 2009). Within each genre-based unit, instruction includes introduction of the genre, discussion and evaluation of good and weak papers during which the rubric is introduced, explanation and think-aloud modeling of the strategies, collaborative writing, guided practice, and further modeling of evaluation in preparation for peer review.

The pedagogical methods for teaching cognitive and metacognitive strategies are complex. Research (Harris et al., 2012; Rietdijk et al., 2017; Traga Philippakos, 2020; Traga Philippakos & Voggt, 2021) has found that professional development on strategy instruction, supported by clear instructional materials and coaching, can support implementation with fidelity and lead to positive effects on student writing. The SSW program includes several supports for fidelity of implementation. First, an instructors' guide includes detailed lesson plans with all needed materials. Second, the PD in advance of instruction applies the principles of strategy instruction; the strategies are explained, modeled, and discussed among instructors; then instructors teach the strategies and receive feedback. Third, instructors receive coaching on fidelity of implementation during the semester.

Prior Research on SSW

The current study follows a prior development project that included two years of design research and a quasi-experimental study. The design research included three iterative rounds of development, implementation, and revision based on quantitative and qualitative analyses (MacArthur & Philippakos, 2012; 2013). The final round of design research included 7 instructors and 13 classes in two levels of developmental writing courses. Pretest and posttest data were available for 114 students (51% male, 51% minority, 20% non-native English

speakers). Significant gains over a semester were found for quality of argumentative essays (*ES* = 1.95) and ratings of grammar/conventions (*ES* = 1.18). In addition, significant improvements were found in self-efficacy (*ES* = 1.55) and affect about writing (*ES* = 0.66) and a significant but smaller increase in mastery goal orientation (*ES* = 0.44). The design research resulted in curriculum materials for a full semester course and plans for professional development (PD).

The quasi-experimental study (MacArthur, Philippakos, & Ianetta, 2015) included 13 instructors and 276 students in two 4-year universities with substantial developmental writing programs. The curriculum was taught for a full semester by 6 treatment instructors while 7 control instructors continued with business-as-usual instruction. Treatment instructors were recruited and matched with others at their institution teaching the same course. Curriculum materials included an instructors' guide with lesson plans and materials and a student book. Treatment instructors received two and a half days of PD prior to the semester and coaching as needed during the semester to support fidelity of instruction. The curriculum was implemented with high fidelity, and significant positive effects were found for overall quality of writing on a persuasive essay (*ES* = 1.22, *p* < .001), and for length (*ES* = .71, *p* < .001), but not for errors in grammar/conventions. Significant positive effects were also found for self-efficacy (*ES* = 0.27, *p* = .006) and mastery motivation (*ES* = 0.29, *p* = .009).

Current Study

The goal of the current study is to provide more rigorous evidence of the effectiveness of the SSW curriculum in a randomized control trial. The SSW curriculum used in the current study follows the instructional approach from the prior study. Following recommendations in a technical report published by IES (Schochet, 2008), we divided the research questions into pre-specified confirmatory and exploratory questions. The primary and secondary questions are

confirmatory; the exploratory questions address other issues of interest including other outcomes and moderating variables.

Primary Research Question

1. In comparison to a randomized business-as-usual control group, do students in treatment classes using the SSW program perform better on overall writing quality as measured by (a) course final exam essays and (b) a standardized writing assessment? Based on the previous quasi-experimental study of SSW (MacArthur, Philippakos, & Ianetta, 2015) and a large body of research on strategy instruction in writing (Graham et al., 2016), we anticipated a large effect on quality of writing on both measures.

Secondary Research Questions

2. Compared to controls, do treatment students exhibit greater motivation, including self-efficacy, mastery goal orientation, beliefs about writing, and affect? As students' knowledge, strategies, and self-regulation of writing improve, and as they see their improvement through self-evaluation, we anticipate that their self-efficacy, engagement in mastery learning, and affect will increase. These hypotheses are supported by the previous design research and quasi-experimental study of SSW (MacArthur, Philippakos, & Ianetta, 2015).

3. Compared to controls, do treatment students perform better on grammar and conventions accuracy as measured by course final exam essays? Grammar instruction, though not a major focus of SSW, is embedded in the revising and editing process as supported by research (Jones et al., 2013). In addition, the focus of revision on clarity of key elements (e.g., thesis, topic sentences) may also contribute to improved sentences. We hypothesize a positive effect of SSW. Prior results have been mixed with gains in the design research

(MacArthur & Philippakos, 2012; 2013) but no significant effect in the quasi-experimental study (MacArthur, Philippakos, & Ianetta, 2015).

Exploratory Questions

4. Compared to controls, do treatment students perform better in reading comprehension as measured by a standardized assessment? The curriculum does not directly teach reading comprehension, but prior research has found that writing instruction has positive impacts on reading (for a review see Graham & Hebert, 2011). We anticipated a small positive effect on reading comprehension.

5. What is the acceptability of the program to instructors and students as measured by interviews? This question addresses instructors' perspectives on the feasibility, effectiveness, and limitations of SSW and students' perspectives on their learning and difficulties. Based on the prior study (MacArthur, Philippakos, & Ianetta, 2015), we anticipated positive evaluations by instructors, but had no specific hypotheses about students.

6. Do the effects of the program differ by gender or race/ethnicity? The National Assessments of Educational Progress (National Center for Education Statistics, 2012) in writing and reading across grades 4, 8, and 12 consistently find higher scores for females than males and for White than Black and Hispanic students. Specifically, in developmental education (Bailey et al., 2010; Chen et al., 2016), Black and Hispanic students are more likely than White students to take developmental writing and less likely to pass developmental courses and FYC; males are less likely to pass developmental writing courses than females. Theoretically, it is possible that groups less likely to pass the course would benefit most from the extra support of SSW; on the other hand, it is also

possible that more successful groups would be more able to take advantage of the

strategies. We did not have any specific hypotheses.

Method

Participants and Sites

Recruitment

Community colleges in the mid-Atlantic region were recruited through emails to college

administrators inviting them to participate in a grant-supported research project on

developmental writing. A three-page attached invitation provided information on the project and

prior research results, the instructional approach, and benefits and requirements for participation.

Requirements included a commitment to participate from 6 to 12 instructors, full-time or adjunct,

with the understanding that instructors would be randomly assigned to treatment (T) or control

(C) conditions and that all instructors would receive a stipend. Treatment instructors would

receive PD and teach using the SSW curriculum for a full semester, while control instructors

would continue with their typical instruction. Colleges were assured that all costs for materials,

PD, and data collection would be covered by the project. Follow-up meetings were arranged with

interested colleges, first with administrators and instructional leaders, and then with instructors.

If administrators were interested, they invited faculty to attend an information session with the

researchers to learn about the study and decide if they wanted to volunteer. The research was

fully explained to instructors to obtain informed consent following procedures approved by the

Institutional Review Board (IRB). To avoid potential contamination of the control group, the

instructional methods were not described in the meeting with instructors, and treatment

instructors were asked during PD not to share materials or discuss the instruction with

colleagues.

Sites and Instructors

Two community colleges from different states participated in the study, college A in the fall semester and college B in the subsequent spring semester. Both colleges were located near large cities and had multiple campuses. College A had campuses located in urban, suburban, and rural areas; college B had campuses in suburban and rural areas. All courses were three credits; courses were eight weeks at college A and 14 weeks at college B but with similar contact hours (32 and 36 hours, respectively). Both colleges offered two levels of developmental writing; students placed in the lower level were expected to take the higher-level course before FYC.

Participants included 19 instructors. Instructors were randomly assigned within college to condition (6 T and 5 C from college A, 4 T and 4 C from college B). At college A, 4 instructors (2 T and 2 C) taught sections of a lower-level developmental writing course; the other 7 at college A and all instructors at college B taught a higher-level developmental writing course. Instructors were blocked by course level for random assignment. Most of the instructors were white women (7 T, 6 C); 4 were white men (3 T, 1 C), and 2 were African American women (2 C). With a few exceptions, all instructors' highest degree was a Masters in English or education. One control instructor had an MS in psychology, and one control instructor had a doctoral degree in education. One treatment instructor had a juris doctorate, and one treatment instructor had a BA in English. Experience teaching college writing ranged from teaching 3 to over 50 college English classes; 4 control instructors and 3 treatment instructors had taught fewer than 10 English courses. Just over half of the instructors were part-time (5 C and 6 T). This information on demographics and experience was gathered using a questionnaire completed by instructors at the time they gave consent.

Students

Students in one course section for each instructor were invited to participate; for instructors teaching more than one section, selection of section was planned to maximize class size and avoid sections scheduled at the same time. A researcher explained the study in each class and obtained informed consent following a script consistent with procedures approved by the IRB. Of the 278 students in the 19 classes, 245 consented to participate in the study (88.1% participation, 93.2% at college A, and 82.4% at college B). Of the 245, 38 students dropped the course or stopped attending, 16 treatment and 22 control students. Overall attrition was 15.5% and the difference by condition (12.6% versus 18.6% in treatment and control groups, respectively) was not significant ($p = .19$). There were no significant demographic differences on gender, age, minority status, or native-English speaking between students who withdrew and students who completed the course (all $p > .10$). With overall attrition of 15.5% and differential attrition of 6.0%, this study sits very close to the threshold of What Works Clearinghouse (WWC) standards for acceptable attrition under "cautious assumptions" (5.9% differential attrition) and well below the standard under "optimistic assumptions" (10.6% differential attrition; see Table II.1, p. 12 in WWC, 2020). Given that attrition in developmental community college courses is typically high (Bailey et al., 2010; Nastal-Dema, 2019), we argue that much of the attrition in our study is unrelated to the intervention, and the WWC standards under optimistic assumptions apply. Furthermore, given that students who drop out of developmental courses are often struggling in the course, potential bias due to greater attrition in the control group would push the mean outcomes for the control group higher, and the treatment effect would be underestimated. In sum, attrition in this study does not exceed WWC thresholds and estimates of SSW impacts are likely conservative.

Overall, 207 students ($n = 96$ control, $n = 111$ treatment) completed the course and are included in the final analysis (see Table 1). Of this sample of 207, 62.1% were female and 37.9% male, 43% White, 38.2% Black, 3.9% Asian, 9.2% Latino, 1% Native American, and 4.8% other; 9.2% of students were born outside of the US, and 11.6% spoke a primary language other than English at home. Demographic information was gathered with a brief questionnaire completed by students when they gave consent to participate. Compared to college A, college B had a larger proportion of male students (($\chi^2(1) = 14.0$, $p < .001$) and students born outside the US (($\chi^2(1) = 23.6$, $p < .001$), but there was no difference in proportion of non-white students (($\chi^2(1) = 1.21$, $p = .271$) . However, there were no significant differences in demographics between treatment and control students: gender ($\chi^2(1) = 0.458$, $p = .50$); minority status ($\chi^2(1) = 0.59$, $p = .44$; born outside US ($\chi^2(1) = 1.90$, $p = .39$).

Measures

Instructor Measures

Treatment instructors were observed to evaluate fidelity of treatment and to provide feedback to enhance fidelity. Control instructors were observed to describe their instruction; they also provided copies of their syllabi and assignments. In addition, treatment instructors were interviewed after the semester.

Fidelity of treatment. Treatment instructors were observed for full class sessions 3-5 times during the semester to rate fidelity of implementation, ranging from 16.7% to 31% of total class time per instructor ($M = 22.7\%$). Observations were scheduled with instructors to observe key lessons in the curriculum (e.g., modeling of strategies, preparation for peer review). The fidelity measure had been developed and applied in prior research (MacArthur, Philippakos, & Ianetta, 2015). Observers took detailed field notes during the class, recording all activities and

capturing as much instructor language and student response as possible and including times at the start of each lesson activity and at least every five minutes. Immediately after class, they edited the notes and then completed the fidelity measure, which included a checklist of lesson components and ratings of quality of key elements of the instructional approach. The checklist included all the lesson activities; each activity was rated 0 to 2 (0 - absent, 1 - done with modifications, 2 - done as in the lesson plan). Quality of implementation was rated on 3-point rubrics for five key elements: introduction of a new genre, analysis of good and weak examples, modeling of the strategy, collaborative practice, and peer review. Only lesson activities and quality elements relevant to the day's lesson were rated.

Observers included three of the authors and three graduate research assistants (RAs) who were thoroughly familiar with the curriculum. Training materials included video recordings of instruction from the design research and field notes from the first experimental study (MacArthur, Philippakos, & Ianetta, 2015). The fidelity measure was explained; the recordings and field notes were discussed and rated together. The field notes were discussed as examples of how to record activities and classroom interactions. RAs then independently rated field notes from six sessions and returned for discussion. On a second set of independent ratings of six sessions, reliability was scored as agreement with the first two authors. For the 3 raters, exact agreements were 84%, 88%, and 98% for components and 75%, 90% and 100% for quality for the three RAs. To prevent rater drift during actual field observations, ratings were checked on 21% of observations, including all six raters and 7 of 9 treatment instructors, either by simultaneous observation by two observers (only two sessions) or by having a second rater score fidelity based on the field notes; exact agreement was 94% on the components checklist and 96% on the quality ratings.

Observations of control instructors. Control instructors were observed three times each for full class sessions (1.5 – 2 hours) by the same staff who observed treatment instructors. Observations were scheduled with instructors and spread across the semester. An observer took detailed field notes, noting for each activity the start and end times, the focus of instruction, as much instructor language and student response as possible, and any assignments or writing tasks. Observers collected copies of instructional materials used in class. After class, observers completed a control observation summary sheet, listing categories of activities with time devoted to each, including explaining genres, explaining writing processes/strategies, conferencing, peer review, collaborative writing, and in-class writing and reading. In addition to the observations, control instructors provided copies of their syllabi and all writing assignments.

Interviews of treatment instructors. Treatment instructors were interviewed after the semester by the researchers to gather information on their perspectives on the curriculum and their students' performance. Semi-structured interviews included questions on instructors' overall evaluation of the curriculum and particular components, impact on student learning and motivation, professional development, and challenges for instructors and students. The interviews were recorded and transcribed by the researchers who did the interviews, only transcribing fully understood complete words; the interviewer checked for accuracy by listening again while reading the transcripts.

Student Measures

At pretest and posttest, students wrote argumentative essays and completed a motivation questionnaire. Standardized scores from reading and writing placement tests prior to the study were obtained from the colleges, and students took the reading test at posttest. At posttest, they also wrote a standardized essay from the National Assessment of Educational Progress. These

measures were completed by all students. In addition, at posttest, a sample of students from the treatment group only was interviewed.

Essay quality, length, and grammar/conventions. In class during the first week of the semester and in the final examination, students wrote argumentative essays without use of sources. Although students learned to write in multiple genres, argument was chosen for the final assessment because of its fundamental importance to academic writing (Wolfe, 2011) as supported by surveys of college composition and English instructors (ACT, Inc., 2012; Milewski et al., 2005). We confirmed that it was taught in the existing courses at both colleges through review of existing syllabi. At each time, students had a choice of two topics that had been evaluated in previous studies (MacArthur & Philippakos, 2013; MacArthur, Philippakos, & Ianetta, 2015) and found to result in essays of similar quality and length (for all comparisons, ESs < 0.20). The writing prompt provided brief directions to take time for planning and revising, to use their own ideas, and to use the word processor and spell checker. Each topic included a brief introduction to the issue, a topic as a question (e.g., Should writing courses be taught online?) and a final direction to take a position and support it with evidence. On both occasions, the directions and prompts were given to students in print and read aloud. The two testing occasions were not comparable. The pretest was written in class in less than an hour and explained to students as a baseline measure of their writing achievement; the posttest was the final examination and students had 90 minutes. The pretest measure was used as a covariate to control for individual differences.

Four RAs independently rated papers for overall quality on a 7-point holistic rubric (1-7) that directed raters to form an overall judgment of quality based on criteria for ideas or content, organization, word choice, sentence fluency, and conventions (see Supplemental Materials). Two

raters scored papers for college A and two others for college B. The separate rating times were necessary to provide scores to college A prior to promised PD for control instructors in the summer. Raters were trained using identical materials; in addition to the rubric, materials included anchor papers and sets of training papers from a prior study using the same topics. Raters were trained to criteria of 70% exact and 90% adjacent agreement on a set of 20 training papers. Agreement within one point is adjacent. Pretest and posttest essays were combined and randomly ordered. All essays were scored by two raters. Any disagreements greater than 1 were resolved through discussion. Scores were averaged for the final score. Interrater reliability was acceptable (Ramineni & Williamson, 2013) with Krippendorff's $\alpha = .77$. We used Krippendorff's α given its demonstrated superiority to other measures of interrater reliability based on ordinal coding of textual documents by multiple raters (Hayes & Krippendorff, 2007).

 Essay length was calculated using the word processor's word-count feature.

 Essays were also scored for errors in grammar and conventions by marking errors and calculating the proportion of T-units that did not have an error. First, essays were divided into T-units and fragments by three raters. A T-unit (Hunt, 1964) is defined as a main clause and all embedded subordinate clauses and phrases; it is a better unit for analysis than the sentence because it is independent of punctuation. Raters were trained to a criterion of 95% agreement on scores for individual T-units ($r = .99$ on total T-units) with consensus scores of two researchers on a set of 30 papers; they then divided the papers for marking. Second, two different raters scored the essays for errors in grammar and conventions. Each T-unit was scored as correct or incorrect, counting fragments as incorrect T-units, and the proportion of correct T-units was calculated. Raters were trained using a scoring manual and essays from prior research with consensus scores from two researchers. Training continued to criteria of correlations with

consensus scores greater than .90 for total errors. Once raters started scoring the actual study

essays, they scored 50 randomly selected essays in common; correlations with the consensus

scores were .96 and .95 for total errors for both raters. The remaining essays were randomly

divided between the two raters for scoring.

Standardized writing assessment. As a distal measure unrelated to the specific

curriculum, students wrote a posttest essay using a retired 12th-grade persuasive prompt from the

National Assessment of Educational Progress (National Assessment of Educational Progress, no

date). Essays were scored for overall quality using the NAEP rubric for persuasive writing,

which is a 6-point scale ranging from 1 = little to no skill, to 4 = adequate, to 6 = effective.

Independent of the project staff, a consultant with expertise in NAEP scoring trained raters

following standard NAEP procedures, which specify training raters to a minimum of 60% exact

agreement (National Center for Education Statistics, 2017). All essays were scored by two raters;

agreement was good with 65% exact and 96% adjacent agreement, and interrater reliability was

also acceptable (Krippendorff's $\alpha = .71$).

Accuplacer Reading assessment. Accuplacer assessments are widely used by

community colleges for placement into developmental courses (Rutschow et al., 2019).

Accuplacer Reading (College Board, 2016) is a computer-adaptive test of reading

comprehension in which students read high school and college level passages and answer

questions about main ideas, inferences, and applications. The developer reports an internal

consistency reliability (Cronbach's alpha) of .87 (Mattern & Packman, 2009). Both colleges used

the Accuplacer Reading assessment, though one of them used a beta version with a different

score range. We obtained pretest and posttest scores from the colleges' testing services. The

pretest scores were part of the colleges' normal assessment for placement. The posttest scores were administered by the colleges' testing centers for the study.

Motivation questionnaire. At pretest and posttest, students completed a motivation questionnaire tapping self-efficacy, goal orientation, beliefs, and affect (MacArthur, Philippakos, & Graham, 2016; Traga Philippakos et al., 2021) The self-efficacy scale included 22 items in three factors: writing tasks and processes (e.g., I can start an essay with an interesting introduction), self-regulation (e.g., I can plan time to get my writing done by the deadline), and grammar (e.g., I can write a paper using correct grammar). Consistent with research recommendations (Pajares & Valiante, 2006), the response scale ranged from 0% to 100% confidence. The goal orientation scale included 14 items assessing mastery goals (e.g., improve how I express my ideas), performance goals (e.g., have my classmates believe I can write well), and avoidance goals (e.g., avoid making mistakes in front of my classmates). The writing beliefs scale with 12 items assessed two factors: beliefs about the importance of substance (e.g., Good writers discover new ideas while writing) and mechanics (e.g., Good writers have to be able to write long sentences correctly). The affect scale included a single factor with 5 items about feelings and attitudes about writing (e.g., The process of writing is satisfying for me). Prior research demonstrated adequate construct validity and internal consistency with Cronbach alphas for factors ranging from .66 to .93; only "beliefs about grammar" was below .70 (Traga Philippakos et al., 2021). Internal consistencies for data in the current study ranged from .73 - .96 except for "beliefs about grammar" at $\alpha = .64$. See Table 4 for reliabilities for the individual sub-scales.

Student Interviews. At posttest, RAs interviewed a sample of students in the treatment condition to understand their perspectives on their course and thoughts about their learning.

Semi-structured interviews included questions on overall evaluation of the course, evaluations of specific components (i.e., strategies, peer review, and SAS), difficulties, and overall learning. Students were randomly selected and invited to participate voluntarily in interviews with the goal of interviewing 3 students per class. Interviews were transcribed either by the RAs interviewers with accuracy checked by a second RA, or by a professional transcription service; only fully understood complete words were transcribed.

Procedures

General Research Procedures

In the first class of the semester, all instructors administered the pretest essay, reading the printed directions and prompts to the students. In the second class, research staff visited all classrooms to explain the study and request informed consent. In that same visit, research staff administered the motivation questionnaire and collected the pretest essays from the instructors as word processor files. Scripted procedures were followed for obtaining consent and administering the motivation questionnaire. Observation schedules were arranged with instructors; treatment observations were planned for key lessons, and control observations were spread throughout the semester. Near the end of the semester, arrangements were made with the colleges' testing centers for students to take the Accuplacer Reading assessment. Students took the assessment in the testing center on their own time before the end of the semester. In the final week of the semester, research staff visited all classes to administer the NAEP writing assessment and the motivation questionnaire. The posttest essay was administered as the final exam for the course, proctored by the research staff and instructors. All writing assessments for the research study (pretest and posttest essays and the NAEP assessment) were completed in classrooms with sufficient computers for all students.

Treatment Instruction

The Supporting Strategic Writers (SSW) curriculum is organized into units of instruction on genres with versions that begin with paragraphs and move to essays or that use essays throughout. At each college, researchers met with instructional leaders to decide which units of instruction to include based on their current course syllabi. At college A in the lower-level course, instruction included narrative, procedural, and argumentative writing proceeding from narrative paragraphs to argumentative essays. The higher-level course developmental course included narrative, cause/effect, and argumentative essays. At college B, instruction began with procedural paragraphs and proceeded to cause/effect and argumentative essays. Students wrote two papers in each genre, the second with less instructor support. The unit on argumentative writing was taught at the end of the semester in all classes.

Each of the genre-based units of instruction followed a common sequence. Instructors and students discussed the purposes for which the genre is useful and the organizational elements. For example, the elements of argument include a controversial issue, the author's position, reasons with evidence, opposing reasons with evidence and rebuttals, and a conclusion that leaves the reader with something to think about. Next, the class discussed a good student example of the genre and applied the evaluation rubric; this was followed by evaluation of a weaker paper.

Then the instructor explained and modeled the writing strategies as applied to the genre. The writing strategies use knowledge about the rhetorical purposes and text structures of academic genres to guide both planning and revising processes. The planning strategy begins with rhetorical analysis of audience, purpose, and genre using the mnemonic TAPFOR (topic, audience, purpose, form, organizational elements of the genre, requirements for assignment). The

audience and purpose determine the genre. Strategies for generating and organizing content include brainstorming and use of a graphic organizer that varies by genre. For example, since arguments are intended to persuade, it is important to consider alternative positions. Thus, for brainstorming, a simple t-chart is used to generate reasons and evidence on both sides, while considering possible audiences. The graphic organizer for argument includes the issue, position and an opposing position, reasons and evidence on both sides, rebuttals for opposing reasons, and a conclusion. Drafting involves following the plan and elaborating on ideas while considering paragraph and sentence construction and cohesive devices like transition words. In the case of argument, students utilize sentence frames to develop the opposing position and rebuttal (e.g., "Some people argue (believe, say, think) that…". Revising is guided by a rubric with evaluation criteria based on the genre elements. For example, the rubric for argumentative writing asks students to find each of their reasons and evaluate whether it is clear, connected to their position, and supported by evidence.

After modeling, the instructor and students engage in collaborative practice, in which students provide the content while the instructor guides them in following the strategy for planning, drafting, and revising. At this point, students begin work on their own essays with support as needed. Once students have written first drafts, the instructor prepares students for peer review by guiding them in collaborative evaluation of papers written by unknown peers, scoring them on the genre-specific rubric and making suggestions for improvement. Students then engage in peer review in pairs, following the same process.

Grammar and conventions are taught in the context of editing. After peer review, instructors provide a brief editing lesson focused on fixing typical errors in the context of complete texts. The curriculum includes a few suggested lessons, but instructors are encouraged

to select the grammar and conventions issues most relevant for their students based on their writing.

Throughout the process, instructors support use of the metacognitive SAS strategies for goal setting, task management, progress monitoring, and reflection. Students write in reflective journals and discuss their journal thoughts in class. Sample prompts for journal entries include "Did I use the planning strategy and did it help?" "What plans have I made for managing when and where to work on my writing?" "What goals will I choose for my next essay?" Teacher feedback using similar questions has been shown to improve student writing (Schunk & Schwartz, 1993). In general, students are taught to apply the four strategies for each writing task – setting goals, managing the task, monitoring use of the strategies and progress, reflecting on their learning, and setting new goals. To support the goal of independent performance, students write a second paper in each genre with less instructor support.

Professional Development

Instructors in the treatment group participated in two or two-and-a-half days of PD in the month before the start of the semester and received in-class coaching on fidelity of implementation during the semester. The half day for college B was cancelled due to weather, and content was covered in two days. Instructors received copies of the Instructor's Guide, which contained detailed lesson plans, and the student book. PD was provided by the primary researchers and was designed to apply the principles of strategy instruction. PD at both colleges followed the same sequence of activities delivered by the same researchers. The first day of PD began with a presentation and discussion of the key principles of SSW, followed by an explanation of the strategies and an overview of the curriculum and teachers' guide. The researchers modeled teaching the strategies from one unit, and the instructors participated as

students, planning and drafting a paper. In small groups, instructors practiced think-aloud modeling of the same planning strategy they had just observed. For homework, the instructors prepared to teach the writing strategies for a genre from another instructional unit using think-aloud modeling, which they did the second day. The researchers provided guidance and feedback with further discussion of the pedagogical methods. Time was also devoted to discussion of the metacognitive strategies. Finally, time was spent discussing the experimental study procedures and requirements.

During the semester, the researchers observed treatment classes 3-5 times to provide feedback on fidelity of implementation. When possible, observers met with instructors after class to provide feedback and to answer questions; in addition, feedback was always shared in writing (Authors), and follow-up questions were addressed either by phone or email. The researchers did not model instruction in instructors' classrooms, and coaching only addressed fidelity of implementation.

Control Instruction

Control instructors continued to use their prior approaches to instruction. The following description of their instruction is based on examination of course syllabi and materials, collection of course assignments, and observations (see Measures). All control syllabi included writing assignments in multiple genres (4 to 6), including an argumentative essay as the final course assignment. Only one instructor required regular writing in a journal. Most of the instructors (7 of 9) were observed devoting some class time to explanation and discussion of genres of writing. About half of the control instructors (5 of 9) taught some strategies for planning. Three provided some instruction on brainstorming, two taught some form of graphic organizer, and two taught outlining. However, only one control instructor modeled the process of using strategies in the

writing process. All instructors gave feedback on student writing and allowed students to submit multiple drafts. However, no peer review was observed in any of the control classes, though two instructors mentioned it on their syllabi. As for metacognitive, self-regulation strategies, none of the control instructors were observed spending time discussing work habits, time management, or goals, and no such information was included in their syllabi. All of the control instructors assigned homework exercises on grammar/conventions and provided individual editing feedback to students. Instructors at college A required grammar assignments and quizzes from an E-textbook, while at college B, assignments varied. Grammar and conventions were not the major focus in any of the control classes. Also, most instructors (8 of 9) allotted some time for students to write and read in class. No collaborative writing by students or teachers with students was observed.

Analysis

Quantitative Analysis

Preliminary analyses checked pretest equivalence between treatment and control groups on demographics (see Participants) and all pretest assessments. For the main analyses, considering that the students were nested within sections taught by different instructors, hierarchical linear modeling (HLM) was used as it allowed the examination of the effects of this nesting (Bryk & Raudenbush, 1992). The effects of the curriculum on the quality, length, and grammatical errors of written responses, the NAEP assessment, and the motivation outcomes were estimated in two-level HLM models with students nested within instructors. College was included in the model as a fixed factor because there were only two colleges and it was used as a blocking variable during random assignment; additionally, we wanted to determine whether the curriculum was differentially effective across colleges with different characteristics. For each

analysis, the relevant pretest score was used as a covariate. Following these main analyses, moderator analyses were conducted to examine whether the treatment was differentially effective by gender and race/ethnicity, and course level (see section on Sites and Instructors); the analyses were re-run with these variables added as factors and included in interactions with the treatment indicator. Finally, a Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) was carried out in order to control the False Discovery Rate (FDR) for the primary impacts across the 14 tested outcomes.

Qualitative Analysis of Interviews

We used directed content analysis (Hsieh & Shannon, 2005) for analysis of instructor interviews. An initial set of codes was developed based on the interview questions, which addressed evaluation of the curriculum and components, impact on student learning and motivation, professional development, and challenges for instructors and students. Two researchers read all the interviews and then independently coded two interviews, adding codes as needed, and discussed the coding and interpretation. Additional interviews were coded independently and discussed until agreement was reached on a full set of codes. Then one researcher took the lead with the second checking the coding and conclusions. Conclusions were checked for disconfirming evidence and variation in perspectives by systematically reviewing the coded interviews. The same process was followed for student interviews using codes based on the student interview questions, including evaluation of the curriculum and components (i.e., writing strategies, SAS, and peer review), how much they had learned, and difficulties experienced.

Results

Preliminary analysis of equivalence between treatment and control groups found no significant differences at pretest for Accuplacer reading scores ($p = .85$, $d = -.09$), baseline essay quality ($p = .26$, $d = -.24$), essay length ($p = .82$, $d = -.04$), or essay grammatical errors ($p = .68$, $d = .06$).[1] However, a significant pretest difference was found for one motivation factor (out of nine), self-efficacy for grammar ($p = .036$, $ES = -.35$), and a nearly significant difference for affect ($p = .066$, $ES = -.29$), both with higher scores for control classes (see Tables 2 & 3).[2] WWC guidelines state that, "Some valid randomization procedures can produce intervention and comparison groups that appear dissimilar based on chance. The WWC does not consider these chance differences to compromise the RCT" (What Works Clearinghouse, 2020, p. 7). However, the WWC does state that baseline differences in an RCT "larger than 0.25 standard deviation for any specified preintervention measure in a domain means that all of the outcomes in the domain fail to satisfy the baseline equivalence requirement" (What Works Clearinghouse, 2020, p. 17). Therefore, this RCT failed to establish baseline equivalence in the "Motivation" domain and

---

[1] At Site B, students in the control condition had significantly higher pretest writing quality scores than treatment students ($p < .05$), but no significant differences were found when data were pooled across the two sites. Given that the main impact analysis requires pooling data across the two sites (i.e., as in our grant proposal), WWC standards do not require any special measures to address baseline imbalance. Furthermore, potential bias due to baseline differences in Site B would likely result in underestimation of the SSW impact.

estimates of the impacts of SSW on motivation outcomes may be biased (i.e., likely underestimated). Pretest scores were used as covariates in all analyses.

Fidelity of Treatment

Fidelity of implementation was measured using a checklist of lesson components and ratings of quality of key elements of strategy instruction (see Measures). Overall fidelity of treatment was good. For implementation of lesson components, mean scores for instructors ranged from 1.24 to 1.90 on a 2-point scale ($M = 1.7$, $SD = 0.25$). For quality of implementation of key features, mean scores for instructors ranged from 1.77 to 3 on a 3-point scale ($M = 2.52$, $SD = 0.44$). Eight of ten instructors had fidelity scores over 1.5 for components and 2.5 for quality; the other two instructors scored below these levels on both measures.

Writing Outcomes

Essay Quality, Length, and Grammar/Conventions

Descriptive data are presented in Table 2, and correlations among measures, including motivation, are in Table 3. For quality, the analysis found a statistically significant effect of condition, $F(1,16) = 29.75$, $p < .0001$, with a very large effect size, Glass's $\Delta = 1.18$, favoring the treatment group. Glass's delta, which uses the posttest control standard deviation, was appropriate because the variance at posttest increased for the treatment group. The covariate pretest quality score was significant ($p = .007$). No main effect was found for college, and no interactions were found between condition and college or course level, indicating that the treatment worked equally well in both colleges and in both course levels. Additionally, the impact estimate remained unchanged after controlling for essay prompt choice for both pretest and posttest.

For length, no significant effects were found for condition, college, or interaction (all *p*s > .35), and the condition by course level interaction was not significant (p > .10). The covariate pretest length was highly significant (*p* < .0001).

On the grammar/conventions measure of proportion of T-units without errors, the analysis found no significant effects for condition, college, or interaction (all *p*s > .25), and the condition by course level interaction was not significant (p > .10). The covariate pretest grammar measure was significant (*p* < .001). Overall performance was low for both groups; on the posttest, mean proportion of t-units without errors was .54 for treatment and .50 for control.

NAEP Assessment

For overall quality on the NAEP assessment, a statistically significant effect of treatment was found, $F(1,16) = 13.42$, $p = .002$, with a moderate to large effect size (Glass's $\Delta = 0.67$) favoring the treatment group. The covariate pretest essay quality was significant (*p* < .0001). No main effects were found for college or the interaction of college and condition, or the interaction of course level and condition.

Moderator Analysis

Analysis of moderator effects of gender and race/ethnicity found no significant main effects (after controlling for the pretest) or interactions with condition for overall essay quality or NAEP assessment.

Motivation

As for the achievement measures, analyses used HLM with students nested within instructors with condition and college as factors and pretest scores on the same measures as covariates. Separate analyses were conducted for nine motivation scales. Data are presented in Table 4.

Self-efficacy

Statistically significant effects favoring the treatment group were found for all three factors: self-efficacy for tasks and processes ($F(1,15) = 21.1$, p $< .001$, $ES = 0.50$), self-efficacy for grammar ($F(1,15) = 11.66$, $p = .004$, $ES = 0.36$) and self-efficacy for self-regulation ($F(1,15) = 11.0$, $p = .004$, $ES = 0.40$).

Affect

A significant positive effect of the treatment was also found on affect ($F(1,15) = 8.83$, $p = .008$, $ES = 0.32$).

Belief

A statistically significant positive effect of treatment was found for beliefs about importance of content ($F(1,15) = 5.6$, $p = .021$, $ES = .29$), but not for beliefs about the importance of mechanics ($p > .15$).

Goal Orientation

No significant effects were found for goal orientation (mastery, performance, avoidance) (all ps $> .35$).

Moderator Analysis

Analysis of moderator effects of gender and race/ethnicity found no significant main effects or interactions with condition for any of the motivation factors.

Accuplacer Reading Posttest

For the Accuplacer reading posttest, only 165 students participated, 94 treatment and 71 control. The tests were administered by the college testing centers, and some students were given the writing test instead of reading; we were unable to correct the error. No significant effect for treatment was found ($p > .5$).

Controls for Multiple Tests

Table 5 presents results from a Benjamini-Hochberg (B-H) procedure to control the False Discovery Rate (FDR; Benjamini & Hochberg, 1995) among the tests of primary impacts for the 14 outcome variables examined. Results show that after restricting the FDR to 10% (i.e., no more than 10% of significant results are expected to be false positives), all of the results deemed significant up to this point remain significant under the B-H procedure. These results hold even if the FDR is set to a very conservative 5%, suggesting that the observed impacts of SSW on seven of our outcome measures are very unlikely to be false positives.

Interviews: Treatment Instructors' Perspectives

All ten instructors gave positive reviews of the curriculum overall and agreed that students benefitted from the experimental curriculum. Curriculum organization, the consistent application of the writing strategy across multiple genres, and a systematic approach to strategy instruction were noted as strengths. All ten mentioned that explicit explanation and modeling of the writing strategies with repetition across the semester were important; together, they provided students with a structured approach to learning the strategies of effective writers. Most (7) also mentioned specific aspects of the strategies, such as the graphic organizers and rubrics, and several (4) noted the alignment of the graphic organizers and rubrics with the elements of the genres. Most (8) instructors found the self-regulatory SAS worthwhile, but a few (3) felt they should have devoted more time to teaching SAS, and a few (2) suggested SAS could be better integrated throughout the curriculum.

When asked to share what was challenging about the curriculum, most (8) teachers indicated that following the curriculum was initially difficult because it meant teaching "differently," out of their "comfort zone." Some instructors (4), all teaching 8-week semesters at

college A, found the pacing challenging. A few (3) reported that peer review was challenging. However, all ten instructors indicated that their confidence grew as they implemented the curriculum, received feedback, and gained understanding of strategy instruction, especially modeling. Many (5) volunteered that the instructional methods were "better" than approaches they had used in other curricula. "It's not just the assignments, but teaching a way of thinking, of approaching the assignments." Some (4) said they were teaching more than usual, and the students responded positively to the accelerated pace. "I liked the structure. I think the structure and pace helps keep the students on task." One instructor shared how positive feedback from students about the modeling encouraged her, "I was surprised how many students came up to me and told me how much they enjoyed it. When you hear about it in PD, you're not sure. It sounds good, but you're not sure how the students are going take it. But so many of them loved it, and they would say – I wish everybody else did this."

When asked directly "if there was anything you think students didn't learn this semester, compared to previous semesters," the only responses were about grammar; three said they hadn't taught as much grammar as they usually did, and one mentioned adding grammar mini-lessons. In contrast, when asked, "How much do you think was learned in comparison to previous semesters?" the consensus among the faculty (10) was that student writing improved with each unit, and that because expectations were higher, and the curriculum more rigorous, students exited the course better prepared for college English.

All ten instructors found the PD valuable, especially the focus on modeling, but some (4) thought more practice might be helpful. "The modeling was probably the hardest thing for me and the students. You need more time to do more modeling." When asked about the in-class observations and feedback, the response from all of the instructors was positive. They found the

feedback timely and described it as "thoughtful," "honest," and "important." One instructor shared the feelings expressed by many, "The feedback was very helpful, and it helped me grow and develop as an instructor."

Interviews: Treatment Students' Perspectives

In all, 27 treatment students were interviewed, 2 to 4 in each class. Nearly all students (26) had highly positive opinions of the course. One student responded negatively to all questions and complained he did not belong in a developmental class. The class experience was described most often as "good, great, excellent" (13) or "helpful" (9). When asked why, students reported the most helpful parts of the class were the teacher's instruction (10), writing strategies (8), graphic organizers (2), and peer review (2). The majority (20) found the planning strategies important and helpful, a few (3) took a neutral stance, while some (4) said they took too long to do, or they didn't plan. One-third (9) found the revising strategy helpful, but an equal number (9) didn't follow it. The majority (18) of students reported positive experiences with the peer review process; they described using genre-specific rubrics to guide peer- and self-evaluation, and giving and receiving feedback from peers, as valuable and helpful. One student shared about peer review, "I thought that it (peer review) was really awesome. I thought it was great to just see what my classmate thought about my paper and to be very open minded to people reading how to change it, and how to improve it, so I thought it really helped." Some students (5) were neutral about peer review and some (4) reported negative experiences (either didn't get feedback, or felt it was rude).

Students were asked to explain the SAS and the responses varied; six described the strategies clearly; ten could briefly name 1 or 2 strategies, and six were not able to describe any, or admitted that they were more focused on learning the writing strategies.

The students gave a range of responses about what was hard in the course; some (6) students said "nothing" was difficult; some (6) said there was "not enough time," and some (5) thought argumentative writing was difficult. One or two individuals mentioned each of the following: developing topics, meeting length requirements, formatting, grammar, editing, and fear. When asked to make curriculum suggestions, most (18) said don't change anything, individuals asked for more grammar, no peer review, and more writing topics.

Two-thirds (18) believed they learned "a lot" about writing in the course, especially about planning and including argument elements, and reported their writing improved "a lot." Three students shared that they had used the strategies in other courses. One had already started FYC, "I'm actually doing English 111. I'm using what I learned here to write another narrative." Another said, "Writing a paper in my history class, I basically wanted to find the reasons it happened, and the evidence that it happened. So, for history it's hard to find evidence and reasons, but I found them, and I basically used the graphic organizer." Many (12) felt more confident in their writing ability or reported grade improvements and attributed these unexpected benefits to learning the writing strategy, using the planning strategies, learning genre-specific elements, using peer review, and teacher feedback.

Discussion

The purpose of this study was to examine the effects of a writing curriculum based on strategy instruction with self-regulation on college developmental writers. To our knowledge, this study was the first randomized experiment on pedagogical methods for writing instruction with such learners. The study found a large positive effect on the quality of student writing ($ES =$ 1.18) on an argumentative essay written as a final examination. The results were supported by positive results on an independent measure of quality of persuasive writing from the NAEP (ES

= 0.67). The results are similar to our earlier quasi-experiment using the same curriculum

(MacArthur, Philippakos, & Ianetta, 2015), which also found positive effects on overall quality

($ES = 1.22$). The findings are also consistent with research on self-regulated strategy

development (SRSD) with elementary and secondary students, which regularly finds effect sizes

over 1.0 (Graham & Perin, 2007; Graham et al., 2016). Thus, the study extends the research on

self-regulated strategy instruction to college developmental courses.

In addition, the study found positive effects of the SSW curriculum on several motivation

measures. Positive effects were found for all three aspects of self-efficacy: self-efficacy for tasks

and processes, for grammar, and for self-regulation. Self-efficacy has consistently been found to

correlate with writing performance (Bruning & Kauffman, 2016; Pajares & Valiente, 2006).

Students with confidence in their ability are more likely to engage with writing tasks and focus

on learning and less likely to be inhibited by anxiety. Self-efficacy is, thus, an important outcome

of instruction, especially for lower achieving writers who often lack confidence (Cox, 2009; Hidi

& Boscolo, 2006). Self-efficacy is developed primarily through experiences of mastery

(Zimmerman & Risemberg, 1997). To provide mastery experiences, the SSW approach provides

systematic strategies to help students approach new writing challenges with supportive

instruction through explanation, modeling, and collaborative practice with gradual release of

responsibility. The metacognitive self-regulation strategies are also intended to support

confidence by helping students to take control of their writing and learning processes. In addition

to efficacy, a positive effect was also found for beliefs about the importance of content in

writing. The SSW approach focuses primarily on content – generating, organizing, and

evaluating it – and students had the opportunity to learn that a focus on content improves quality.

A positive effect on affect was also found. Positive effects on self-efficacy and mastery

motivation, and a nearly significant effect on affect (p = .067) were also found in the previous quasi-experimental study (MacArthur, Philippakos, & Ianetta, 2015). These motivational gains may support students' performance in subsequent courses.

We argue that the key principles, or components, of strategy instruction with self-regulation work together to enhance the success of students in developmental courses, increasing both writing achievement and motivation. Scholars in basic writing (Shaughnessy, 1977; Rose, 1989; Bartholomae, 2005) describe their students as underprepared for the demands of college writing rather than lacking in capacity. Facing new and demanding tasks, they do not feel in control of the writing process. SSW integrates three key components to provide the needed support: strategies based on rhetorical analysis and genres, self-regulation strategies to guide learning, and pedagogical methods such as think-aloud modeling and peer review. The genre-based writing strategies are based on the rhetorical purposes, text structures, and linguistic features of genres – common concepts in college English. They provide an initial map for students unsure about how to engage in the writing process. At the same time, the strategies can be flexibly adapted to many writing tasks through rhetorical analysis and genre, and the curriculum includes multiple genres. The self-regulation strategies – goal setting, task management, progress monitoring, and reflection – focus directly on increasing ownership and control of writing and learning processes. Through journal writing and class discussion about these self-regulation strategies, students are encouraged to reflect on how the writing strategies work for them and to understand that they can control their own learning. The pedagogical methods are critical. Strategies are explained clearly and think-aloud modeling makes the cognitive processes visible and available for class discussion. The emphasis on learning evaluation criteria through model essays and peer review contributes to knowledge about the

characteristics of quality writing; self-evaluation is also an important aspect of self-regulation. As students' knowledge, strategies, and sense of control over the writing process grow, their writing performance and their motivation to write improve.

In general, instructors implemented the curriculum with high fidelity. They responded positively to the strategies and pedagogical methods, while recognizing the need for PD to develop proficiency in methods new to them, such as think-aloud modeling, collaborative writing, and the SAS. Evidence from observations, interviews, and outcomes indicates that the curriculum is feasible, and the PD was supportive of instructors' implementation. For larger scale dissemination, methods for providing PD more efficiently might be needed. Students also found SSW helpful overall and thought that their writing had improved, and many also reported increased confidence in their writing. Future research could examine how students transfer strategies to other courses and the modifications they make.

Moving on to the exploratory questions, the study did not find statistically significant effects on errors in grammar and conventions. Although the SSW curriculum did not include a major focus on grammar, it did follow recommended practice in embedding grammar instruction and instructor feedback in the revising process (Jones et al., 2013). Control instructors all assigned grammar homework exercises in addition to lessons and feedback, but grammar instruction was not a major focus in any of the control classes. Both treatment and control students had significant problems with grammar and conventions, making errors on the posttest essay, in nearly half of their sentences. The prior study (MacArthur, Philippakos, & Ianetta, 2015) also found no effects of the curriculum on grammar.

In addition, no significant effects were found on reading comprehension. Although the SSW curriculum did not include explicit instruction in reading, substantial research has found

that writing instruction has moderately large effects on reading outcomes (for a meta-analysis, see Graham & Hebert, 2011). Also, the focus on the text structure of genres in SSW is consistent with reading comprehension interventions focused on text structure, which have found positive effects (Hebert et al., 2016). However, the meta-analysis of Hebert et al. (2016) found small and inconsistent effects ($ES = .13$; 4 of 8 studies with significant effects) for outcomes measured with standardized comprehension assessments. Thus, one possible explanation of null findings in this study is the use of the standardized Accuplacer reading assessment, which was developed for purposes of placing students into developmental classes rather than as an outcome measure. The decision to use the measure was made to avoid further class time spent on testing, but future research should use additional reading measures.

The study also explored potential differential effects by student subgroups. No interaction effects were found for gender or race/ethnicity, indicating that the treatment worked equally well for both groups.

Limitations and Implications for Future Research

Although we found significant differences for motivation on 5 of 9 scales, 2 of the scales had pretest differences greater than .25 *SD;* therefore, the study did not meet WWC standards for baseline equivalence (What Works Clearinghouse, 2020, p. 17). Since the pretest differences favored the control group, any bias would likely underestimate the effects. Furthermore, the positive effects on self-efficacy were consistent with prior research on SSW (MacArthur, Philippakos, & Ianetta, 2015). Nonetheless, the results should be interpreted with caution.

One area for future research is instructional methods for improving grammar and conventions. Grammar instruction is a controversial topic in college writing. On one hand, nearly all textbooks marketed for developmental writing courses include substantial sections on

grammar instruction (MacArthur, Philippakos, & Ianetta, 2015). On the other hand, research has found that traditional grammar instruction has minimal impact on student writing (Hillocks, 1984; Graham & Perin, 2007). In addition, grammar instruction should be sensitive to students' right to their own language (Kinloch, 2005). Some research has found modest positive effects of sentence combining instruction (Andrews et al., 2006) and contextualized grammar instruction focused on using syntactic structures to convey meaning more precisely (Jones et al., 2013). Future research should address this important instructional outcome and how to integrate grammar instruction with other instruction. The issue of grammar instruction may be especially important for second-language (L2) learners, who are more likely to need and benefit from such instruction (Ferris, 2012). Future research could evaluate adaptations of the instructional approach with L2 learners who attend community college developmental courses.

Further, in this study, instructors helped students to develop metacognitive, self-regulatory strategies. Often, such content is covered in separate study-skills courses. Future research could examine the effects of integrating metacognitive strategies into other developmental and first-year courses. One experimental study found positive effects of instruction in self-regulation in developmental math courses (Zimmerman et al., 2011); more research could be conducted.

An important area of needed research is on integrated approaches to reading and writing. instruction. The current study focused on writing based on students' background knowledge without use of any sources or explicit integration of reading and writing. However, for success in college, students need to read source materials critically, analyze, summarize, and synthesize information across sources to write their own essays (Council of Writing Program Administrators, 2014). Currently, many colleges offer integrated reading and writing courses.

Further research is needed on instructional methods for those courses as well as for writing courses that focus on writing using sources. We have developed and are currently investigating an expanded curriculum that includes strategies for critical reading of source texts, note-taking and summarization, and writing essays using sources.

Conclusion

Reform policies and practices in developmental education have emphasized structural changes in placement and organization of courses. Research has provided support for use of compressed or combined courses (Hodara & Jaggars, 2014) and co-requisite courses (Cho et al., 2012; Daugherty et al., 2018; Miller et al., 2020), finding positive effects on enrollment in FYC and other academic outcomes. However, little research has investigated changes in instructional content or methods; more research should be conducted on instructional methods for college writing, in particular for underprepared students. We argue that such research should incorporate strategy instruction principles that have been shown consistently to improve students' writing performance and confidence. The principles of systematic strategies based on genres, instruction in self-regulation strategies, and pedagogical methods such as think-aloud modeling might prove beneficial in other instruction in developmental courses. The current study demonstrates that such research has potential as part of the solution to concerns about the success of underprepared students.

References

Abbott, S. P., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology, 102*, 281-298. https://doi.org/10.1037/a0019318

ACT, I. (2012). *ACT National Curriculum Survey 2012: English Language Arts.* Retrieved from Iowa City: IA: http://www.act.org/research/policymakers/pdf/NCS-EnglishLangArts.pdf

Adams, P., Gearhart, S., Miller, R., & Roberts, A. (2009). The Accelerated Learning Program: Throwing open the gates. *Journal of Basic Writing, 28*(2), 50-69. Retrieved from http://www.jstor.org/stable/43443881

Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., Robinson, A., & Zhu, D. (2006). The effect of grammar teaching on writing development. *British Educational Research Journal, 32*, 39-55. https://doi.org/10.1080/01411920500401997

Attewell, P., Lavin, D., Domina, T., & Levey, T. (2006). New evidence on college remediation. *Journal of Higher Education, 77*(5), 886-924. https://www.tandfonline.com/doi/abs/10.1080/00221546.2006.11778948?journalCode=uhej20

Bailey, T., Bashford, J., Boatman, A., Squires, J., Weiss, M., Doyle, W., Valentine, J. C., LaSota, R., Polanin, J. R., Spinney, E., Wilson, W., Yeide, M., & Young, S. H. (2016). *Strategies for postsecondary students in developmental education: A practice guide for college and university administrators, advisors, and faculty*. Washington, DC: Institute of Education Sciences, What Works Clearinghouse.

Bailey, T., Jeong, D. W., & Cho, S.-W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review, 29*, 255-270. https://doi.org/10.1016/j.econedurev.2009.09.002

Bartholomae, D. (2005). *Writing on the margins: Essays on composition and teaching*. New York: Palgrave MacMillan. https://doi.org/10.1007/978-1-4039-8439-5

Bazerman, C. (2016). What do sociocultural studies of writing tell us about learning to write? In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research, 2nd ed.* (pp. 11-23). New York: Guilford. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300. Retrieved from https://www.jstor.org/stable/2346101

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

Bickerstaff, S., Barragan, M., & Rucks-Ahidiana, Z. (2017). Experiences of earned success: Community college students' shifts in college confidence. *International Journal of Teaching and Learning in Higher Education, 29*, 501-510.

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology, 105*, 25-38. https://doi.org/10.1037/a0029692

Bruning, R., & Kauffman, D. F. (2016). Self-efficacy beliefs and motivation in writing development. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research, 2nd ed.* (pp. 160-173). New York: Guilford.

Bryk, A. S., & Raudenbush, S. W. (1992). *Advanced qualitative techniques in the social sciences, 1.Hierarchical linear models: Applications and data analysis methods.* New York: Sage Publications, Inc.

Chen, X. (2016). *Remedial coursetaking at U.S. public 2- and 4-year institutions: Scope, experiences, and outcomes.* NCES report 2016-405. Washington, DC: National Center for Educational Statistics, U.S. Department of Education. Retrieved from Washington, DC: http://nces.ed.gov/pubsearch

Chen, X., Elliott, B.G., Kinney, S.K., Cooney, D., Pretlow, J., Bryan, M., Wu, J., Ramirez, N.A., and Campbell, T. (2019). *Persistence, Retention, and Attainment of 2011–12 First-Time Beginning Postsecondary Students as of Spring 2017 (First Look)* (NCES 2019-401). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved 2/1/2020 from https://nces.ed.gov/pubsearch. https://doi.org/10.1037/e492162006-002

Cho, S.-W., Kopko, E., Jenkins, D., & Jaggars, S. S. (2012). New evidence of success for community college remedial English students: Tracking the outcomes of students in the Accelerated Learning Program (ALP). (CCRC Working Paper No. 53). New York, NY: Columbia University, Teachers College, Community College Research Center.

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology, 103*, 73-84. https://doi-org.udel.idm.oclc.org/10.1037/a0021950

College Board. (2016, Oct.). *Accuplacer program manual*. New York: College Board. https://accuplacer.collegeboard.org/accuplacer/pdf/accuplacer-program-manual.pdf
https://doi.org/10.1177/0091552109338390

Council of Writing Program Administrators. (2014). WPA outcomes statement for first-year

composition (v3.0). *Journal of the Council of Writing Program Administrators, 38*(1),

144-148. https://www.jstor.org/stable/378996?seq=1#metadata_info_tab_contents

Cox, R. D. (2009). "It was just that I was afraid": Promoting success by addressing students' fear

of failure. *Community College Review, 37*, 52-80.

https://doi.org/10.1177/0091552109338390

Daugherty, L., Gomez, C. J., Carew, D. G., Mendoza-Graf, A., & Miller, T. (2018). Designing

and implementing corequisite models of developmental education: Findings from Texas

community colleges. Santa Monica, CA: RAND, Inc. https://doi.org/10.7249/rr2337

Dole, J., Nokes, J., & Drits, D. I. (2009). Cognitive strategy instruction: Past and future. . In S. E.

Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension.* (pp. 347-

372). New York: Routledge.

Duke, N. K., & Pearson, P. D. (2009). Effective practices for developing reading comprehension.

The Journal of Education, 189(1/2), 107-122. https://www.jstor.org/stable/42748663

Edgecombe, N., Jaggars, S. S., Xu, D., & Barragan, M. (2014). *Accelerating the integrated*

instruction of developmental reading and writing at Chabot College. CCRC Working

Paper #71. New York: Community College Research Center, Teachers College.

Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., & Stevens, D. D. (1991).

Making writing strategies and self-talk visible: Cognitive strategy instruction in writing

in regular and special education classrooms. *American Educational Research Journal,*

28, 337-372. https://doi.org/10.3102/00028312028002337

Ferris, D. R. (2012). Written corrective feedback in second language acquisition and writing

studies. *Language Teaching Research, 45*, 446-459.

https://doi.org/10.1017/S0261444812000250

Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research, 57*, 481-

506. https://doi.org/10.3102/00346543057004481

Fitzgerald, J., & Shanahan, T. (2000). Reading and writing relations and their development.

Educational Psychologist, 35(1), 39-50. https://doi.org/10.1207/s15326985ep3501_5

Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational

Psychologist, 54,* 258-279. https://doi:10.1080/00461520.2018.1481406

Graham, S., & Harris, K. R. (2012). The role of strategies, knowledge, skill, and will in a 30-year

program of writing research (with homage to Hayes, Fayol, and Boscolo). In V. W.

Berninger (Ed.), *Past, Present, and Future Contributions of Cognitive Writing Research

to Cognitive Psychology* (pp. 177-196). New York: Psychology Press.

https://doi.org/10.4324/9780203805312

Graham, S., Harris, K., & Chambers, A. B. (2016). Evidence-based practice and writing

instruction: A review of reviews. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.),

Handbook of writing research (2nd ed., pp. 211-226). New York: Guilford.

https://doi.org/10.1007/s11145-008-9121-2

Graham, S., Harris, K. R., & Hebert, M. (2011). It is more than just the message: Presentation

effects in scoring writing. *Focus on Exceptional Children, 44*(4), 1-13.

https://doi.org/10.17161/fec.v44i4.6687

Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 81*, 710-744. https://doi.org/10.17763/haer.81.4.t2k0m13756113566

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

Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., Barkel, A., Kavanaugh, C., & Talukdar, J. (2018). Reading for writing: A meta-analysis of the impact of reading interventions on writing. *Review of Educational Research, 88*(2), 243-284. https://doi-org.udel.idm.oclc.org/10.3102/0034654317746927

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*, 879-896. https://doi.org/10.1037/a0029185

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

Harris, K. R., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. *British Journal of Educational Psychology Monograph Series II, 6*, 113-135. https://doi.org/10.1348/978185409x422542

Harris, K. R., Graham, S., MacArthur, C. A., & Santangelo, T. (2018). Self-regulation and writing. In B. Zimmerman and D.H. Schunk (Eds.), *Handbook of self-regulation of learning and performance, 2nd ed*. (pp. 138-152). New York: Routledge.

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

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C.

    M. Levy & S. Ransdell (Eds.), *The Science of Writing* (pp. 1-27). Mahwah, NJ: Erlbaum.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure

    for coding data. *Communication Methods and Measures, 1*(1), 77-89.

    https://doi.org/10.1080/19312450709336664

Hebert, M., Bohaty, J. J., Nelson, J. R., & Brown, J. (2016). The effects of text structure

    instruction on expository reading comprehension: A meta-analysis. *Journal of*

    Educational Psychology, 108(5), 609-629.

    https://doi.org.udel.idm.oclc.org/10.1037/edu0000082

Hidi, S., & Boscolo, P. (2006). Motivation and writing. In C. A. MacArthur, S. Graham, & J.

    Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 144-157). New York: Guilford.

Hillocks, G. (1984). What works in teaching composition: A meta-analysis of experimental

    treatment studies. *American Journal of Education, 93*, 133-170.

    https://doi.org/10.1086/443789

Hodara, M., & Jaggars, S. S. (2014). An examination of the impact of accelerating community

    college students' progression through developmental education. *The Journal of Higher*

    Education, 85(2), 246-276. https://doi.org/10.1353/jhe.2014.0006

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis.

    Qualitative Health Research, 15(9), 1277-1288.

    https://doi.org/10.1177/1049732305276687

Hunt, K. W. (1964). *Differences in grammatical structures written at three grade levels, the*

    structures to be analyzed by transformational methods. Tallahassee: Florida State

    University.

Jones, S., Myhill, D., & Bailey, T. (2013). Grammar for Writing? An Investigation of the Effects of Contextualised Grammar Teaching on Students' Writing. *Reading and Writing: An Interdisciplinary Journal, 26*, 1241-1263. https://doi.org/10.1007/s11145-012-9416-1

Kauffman, D. F., Zhao, R., Dempsey, M. S., Zeleny, M. G., Wang, S., & Bruning, R. H. (2010). *Achievement goals' impact on 11th grade students' affect, self-efficacy, and writing achievement.* Paper presented at the American Educational Research Association annual meeting, Denver, CO. https://doi.org/10.3389/fpsyg.2017.01406

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing, 18*, 30-43. https://doi.org/10.1016/j.jslw.2008.06.002

MacArthur, C. A. (2011). Strategies instruction. In K. R. Harris, S. Graham, & T. Urdan (Eds*.), Educational psychology handbook: Vol. 3. Application to learning and teaching* (pp. 379-401). Washington, DC: American Psychological Association.

MacArthur, C. A. (2016). Instruction in evaluation and revision. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 272-287). New York: Guilford.

MacArthur, C. A., & Graham, S. (2016). Writing research from a cognitive perspective. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 24-40). New York: Guilford.

MacArthur, C. A., & Philippakos, Z. A. (2012). Strategy instruction with college basic writers: A design study. In Gelati, C., Arfé, B., & Mason, L. (Eds.), *Issues in writing research* (pp. 87-106). Padova: CLEUP.

MacArthur, C. A., & Philippakos, Z. A. (2013). Self-regulated strategy instruction in developmental writing: A design research project. *Community College Review*, 41, 176-195. https://doi.org/10.1177/0091552113484580

MacArthur, C. A., Philippakos, Z. A., & Graham, S. (2015). A multi-component measure of writing motivation with basic college writers. *Learning Disability Quarterly*, *39,* 31-43. https://doi.org/10.1177/0731948715583115

MacArthur, C. A., Philippakos, Z. A., & Ianetta, M. (2015). Self-regulated strategy instruction in college developmental writing. *Journal of Educational Psychology*. 107, 855-867. https://doi.org/10.1037/edu0000011

MacArthur, C. A. (Principal Investigator). (2016-2020). *Supporting Strategic Writers: Effects of an innovative developmental writing program on writing and reading outcomes (*Grant R305A160242). Institute for Education Sciences, U. S. Department of Education.

Mattern, K. D., & Packman, S. (2009). Predictive validity of ACCUPLACER scores for course placement: A meta-analysis (Research Report No. 2009-2). New York, NY: College Board. Retrieved from https://eric.ed.gov/?id=ED561046

McCutchen, D. (2006). Cognitive factors in the development of children's writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 115-130). New York: Guilford.

Milewski, G. B., Johnson, D., Glazer, N., & Kubota, M. (2005). *A survey to evaluate the alignment of the new SAT Writing and Critical Reading sections to curricula and instructional practices*. New York, NY: College Entrance Examination Board. https://doi.org/10.1002/j.2333-8504.2005.tb01984.x

Miller, T., Daugherty, L., Martorell, P., & Gerber, R. (2020). *Assessing the effect of corequisite English instruction using a randomized controlled trial*. Washington, DC: American Institutes for Research. https://collegecompletionnetwork.org/sites/default/files/2020-05/ExpermntlEvidncCoreqRemed-508.pdf

Nastal-Dema, J. (2019). Beyond tradition: Fairness, placement, and success at a two-year college. *Journal of Writing Assessment, 12*, 1-18. http://journalofwritingassessment.org/article.php?article=136

National Assessment of Educational Progress. (no date). *NAEP Questions Tool, Persuasive writing, grade 12, 2011*. Retrieved from https://nces.ed.gov/NationsReportCard/nqt

National Center for Education Statistics. (2012). *The nation's report card: Writing 2011* (NCES 2012–470). Institute of Education Sciences, U.S. Department of Education, Washington, D.C. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf

National Center for Education Statistics. (2017, May). *NAEP Technical Documentation Within-Year Interrater Agreement*. Retrieved from https://nces.ed.gov/nationsreportcard/tdw/scoring/scoring_within.aspx

National Center for Education Statistics. (2012). *The Nation's Report Card: Writing 2011*. Retrieved from Washington, DC: https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf

Oyserman, D. (2007). Social identity and self-regulation. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social Psychology: Handbook of Basic Principles (second edition)* (pp. 432-453). New York: Guilford Press.

Pajares, F., Johnson, M., & Usher, E. (2007). Sources of self-efficacy beliefs of elementary, middle, and high school students. *Research in the Teaching of English, 42*, 104-120. https://www.jstor.org/stable/40171749?seq=1#metadata_info_tab_contents

Pajares, F., & Valiante, G. (2006). Self-efficacy beliefs and motivation in writing development. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 158-170). New York: Guilford.

Pearson, P. D., & Gallagher, M. C. (1983b). The instruction of reading comprehension. *Contemporary Educational Psychology, 8*(3), 317–344. https://doi.org/10.1016/0361-476x(83)90019-x

Perin, D. (2013). Literacy skills among academically underprepared students. *Community College Review, 41,* 118-136. https://doi.org/10.1177/0091552113484057

Perin, D., & Holschuh, J. P. (2019). Teaching academically underprepared postsecondary students. *Review of Research in Education, 43*, 363-393. https://doi.org/10.3102/0091732x18821114

Philippakos, Z.A., & MacArthur, C.A. (2016). The effects of giving feedback on the persuasive writing of fourth and fifth-grade students. *Reading Research Quarterly, 51,* 419–433. https://doi.org/10.1002/rrq.149

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*, 544-555. https://doi.org/10.1037/0022-0663.92.3.544

Ramineni, C., & Williamson, D. M. (2013). Automated essay scoring: Psychometric guidelines and practices. *Assessing Writing, 18*, 25-39. https://doi.org/10.1016/j.asw.2012.10.004

Rietdijk, S., Janssen, T., van Weijen, D., & van den Bergh, H. (2017). Improving writing in primary schools through a comprehensive writing program. *Journal of Writing Research, 9*(2), 173-225. https://doi.org/10.17239/jowr-2017.09.02.04

Rose, M. (1989). *Lives on the boundary: The struggles and achievements of America's underprepared*. New York: Free Press.

Rutschow, E. Z., Cormier, M. S., Dukes, D., & Cruz Zamora, D. (2019, Nov.). *The changing landscape of developmental education practices: Findings from a national survey and interviews with postsecondary institutions.* New York: MDRC and Community College Research Center, Teachers College.

Santangelo, T., Harris, K., & Graham, S. (2016). Self-regulation and writing: Meta-analysis of the self-regulation processes in Zimmerman and Risemberg's model. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research, 2nd ed.* (pp. 174-193). New York: Guilford.

Schochet, P. Z. (2008). *Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations*. Retrieved from Washington, DC: : http://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=NCEE20084018

Shaughnessy, M. P. (1977). *Errors and expectations: A guide for the teacher of basic writing*. Oxford University Press.

Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology, 18*, 337-354. https://doi.org/10.1006/ceps.1993.1024

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-
regulation of reading and writing through modeling. *Reading & Writing Quarterly, 23*, 7-
25. https://doi.org/10.1080/10573560600837578

Shanahan, T. (2016). Relations between reading and writing development. In C. A. MacArthur,
S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (2nd ed., pp. 194-207).
New York: Guilford.

Snyder, T.D., de Brey, C., and Dillow, S.A. (2019). Digest of Education Statistics 2018 (NCES
2020-009). National Center for Education Statistics, Institute of Education Sciences, U.S.
Department of Education. Washington, DC. https://nces.ed.gov/programs/digest/d18/

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of
Educational Research, 68*, 249-276. https://doi-rg.udel.idm.oclc.org/10.2307/1170598

Traga Philippakos, Z. A. (2020). A yearlong, professional development model on genre-based
strategy instruction on writing. *The Journal of Educational Research. 113*(3), 177-190.
https://doi.org/10.1080/00220671.2020.1767531

Traga Philippakos, Z. A., & Voggt, A. (2021). The effects of distant professional development
model on second-grade teachers' instruction and the quality of procedural papers.
Reading and Writing: An Interdisciplinary Journal. Advance online publication.
https://doi.org/10.1007/s11145-021-10120-1

Traga-Philippakos, Z., Wang, C., & MacArthur, C. A. (2021, July). *Writing motivation of college
students in basic writing and first-year composition classes: Confirmatory factor analysis
of a scale on goals, self-efficacy, beliefs, and affect*. Paper presented at the annual
conference of the Society for the Scientific Study of Reading.

What Works Clearinghouse. (2020). What Works Clearinghouse Standards Handbook, Version 4.1. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. This report is available on the WWC website at https://ies.ed.gov/ncee/wwc/handbooks. https://doi.org/10.1037/e301462004-001

Wolfe, C. R. (2011). Argumentation across the curriculum. *Written Communication, 28*, 193-219. https://doi.org/10.1177/0741088311399236

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling, 53*(1), 141-160.

Zimmerman, B., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology, 22*(73-101), 73-101. https://doi.org/10.1006/ceps.1997.0919

Table 1

Demographic Information and Reading Pretest Scores of Student Participants

	Site A		Site B		Total	
	Treatment	Control	Treatment	Control	Treatment	Control
	n = 66	n = 60	n = 45	n = 36	n = 111	*n* = 96
Gender						
Female	72.7%	71.7%	40.9%	52.8%	60.0%	64.6%
Ethnicity						
White	47.0%	45.0%	31.1%	47.2%	40.5%	45.8%
Black	31.8%	31.7%	57.8%	36.1%	42.3%	33.3%
Asian	3.0%	3.3%	6.7%	2.8%	4.5%	3.1%
Latino	13.6%	11.7%	2.2%	5.6%	9.0%	9.4%
Native						
American	1.5%	1.7%	0.0%	0.0%	0.9%	1.0%
Other	3.0%	6.7%	2.2%	8.3%	2.7%	7.3%
Age	25.4	24.4	24.8	25.1	25.1	24.7
M (SD)	(8.9)	(14.1)	(6.9)	(8.0)	(8.1)	(11.7)
Native English						
speakers	86.4%	91.7%	88.9%	86.1%	87.4%	89.6%
Born in the						
USA	97.0%	91.7%	84.4%	86.1%	91.9%	89.6%
Accuplacer						
reading						
pretest[ab]	138.21	136.83	57.29	58.45		
M (SD)	(11.14)	(11.22)	(19.87)	(19.24)	NA	NA

Writing						
quality pretest	2.52	2.52	2.23	2.76	2.40	2.61
M (SD)	(0.98)	(0.73)	(0.76)	(0.79)	(0.91)	(0.76)

[a] Two separate versions of the Accuplacer Reading (College Board, 2016) were administered at Site A and Site B.

[b] Sample size varied as follows: For site A, treatment $n = 58$, control $n = 53$; for site B, treatment $n = 45$, control $n = 33$.

Table 2

Correlations Among Study Outcome Measures

	Outcome Measure	A	B	C	D	E	F	G	H	I	J
A	Essay Quality	——									
B	Essay Length	.44***	——								
C	Grammar	.12~	.12~	——							
D	NAEP Quality	.49***	.39***	.22**	——						
E	Goals Mastery	.09	-.08	.08	.09	——					
F	Goals Performance	-.04	.05	-.02	-.02	.18*	——				
G	Goals Avoidance	-.08	-.09	.01	-.03	.03	.44***	——			
H	Self-Efficacy Task/Process	.29***	.25***	.13~	.31***	.19**	.16*	-.08	——		
I	Self-Efficacy Grammar	.18*	.15*	.14*	.23**	.11	.13~	-.10	.78***	——	
J	Self-Efficacy Self-Regulation	.23**	.13~	.12~	.18**	.22**	.08	-.02	.81***	.70***	—
K	Beliefs Substance	.07	-.04	-.04	.03	.36***	.08	.00	.34***	.20**	.41
L	Beliefs Mechanics	-.14~	-.11	-.11	-.20**	-.02	.19**	.13~	-.03	.01	.06
M	Affect	.11	.07	.06	.05	.20**	-.01	-.02	.36***	.21**	.47

***p<.001, **p<.01, *p<.05, ~p<.10

Note: Essay quality was rated on a 7-point rubric. Length is the total number of words. Grammar is percent of T-units without errors. NAEP quality was rated on the 6-point NAEP rubric. The motivation scores include 3 goal orientation scores (mastery, performance, and avoidance), 3 self-efficacy scores (tasks/processes, grammar. self-regulation), 2 beliefs scores (importance of substance and mechanics), and 1 affect score.

Table 3

Achievement Outcomes: Essay Quality, Length, and Grammar; NAEP Writing Quality;

  Accuplacer Reading

		Treatment M (SD)	Control M (SD)
Essay Quality	Pretest	2.4 (0.91)	2.6 (0.76)
	Posttest	4.4 (1.0)	3.1 (1.1)
	Adjusted post	4.35 ***	3.01
Essay Length	Pretest	237 (114)	241 (94)
	Posttest	471 (127)	436 (174)
	Adjusted post	472	436
Essay Grammar	Pretest	.46 (.21)	.44 (.24)
	Posttest	.54 (.19)	.50 (.20)
	Adjusted post	.543	.509
NAEP Quality	Posttest only	3.21 (0.94)	2.76 (0.84)
	Adjusted post[b]	3.28**	2.72
Accuplacer Reading[a]	Pretest	102.9 (43.2)	106.8 (41.1)
	Posttest	112.7 (28.3)	118.0 (36.0)
	Adjusted post	104.3	104.6

Note: Essay quality was rated on a 7-point rubric. Length is the total number of words. Grammar

is percent of T-units without errors. NAEP quality was rated on the 6-point NAEP rubric.

[a] n = 165 (94 treatment; 71 control); [b] covariate pretest essay quality

*** $p < .001$; ** $p < .01$; Quality ES (Glass Δ) = 1.75; NAEP Quality ES (Glass Δ) = 0.67

Table 4

Motivation Outcomes: Self-Efficacy, Beliefs about Writing, and Affect

		Treatment			Control		
	Cronbach alpha	Pre	Post	Adjusted post	Pre	Post	Adjusted post
Self-efficacy (SE)	.96						
SE Task/ process	.94	60.1 (17.9)	76.1 (15.1)	77.5***	63.9 (17.2)	69.3 (17.3)	68.9
SE grammar	.89	50.6 (18.8)	65.4 (17.2)	67.5**	57.5 (20.3)	63.4 (65.4)	60.54
SE self- regulation	.78	64.5 (16.8)	76.1 (13.9)	77.4**	68.7 (15.9)	71.7 (16.7)	70.0
Beliefs							
Substance	.85	3.8 (0.6)	4.2 (1.0)	4.15*	4.0 (0.6)	4.0 (0.7)	3.95
Mechanics	.64	2.7 (0.7)	2.6 (0.9)	2.67	2.8 (0.8)	2.8 (0.8)	2.86
Goals							
Mastery	.81	4.07 (0.67)	4.06 (0.67)	4.23	4.00 (0.67)	4.02 (0.67)	4.24

Performance	.80	2.80	2.93	2.93	2.82	3.04	3.05
		(1.03)	(1.19)		(1.09)	(1.14)	
Avoidance	.83	2.74	2.63	2.57	2.92	2.80	2.71
		(1.13)	(1.11)		(1.04)	(0.93)	
Affect	.87	3.1	3.3	3.42**	3.3	3.2	3.12
		(0.9)	(0.8)		(0.8)	(0.9)	

*** $p < .001$; ** $p < .01$; ** $p < .05$

Table 5

Benjamini-Hochberg Results Restricting FDR to 10%

Outcome Variable	p-value	Rank	B-H Critical Value	Significant?
Quality	.0001	1	.0071	Yes
SE Task/process	.0003	2	.0143	Yes
NAEP Quality	.0021	3	.0214	Yes
SE self-regulation	.0043	4	.0286	Yes
SE grammar	.0046	5	.0357	Yes
Affect	.0077	6	.0429	Yes
Beliefs Substance	.0213	7	.0500	Yes
Beliefs Mechanics	.1062	8	.0571	No
Grammar	.2500	9	.0643	No
Length	.3722	10	.0714	No
Goals Avoidance	.3787	11	.0786	No
Goals Performance	.4648	12	.0857	No
Accuplacer Reading	.7863	13	.0929	No
Goals Mastery	.8912	14	.1000	No