Working Paper: Impact of the National Writing Project’s
College-Ready Writers Program in High-Need Rural Districts

H. Alix Gallagher
Nicole Arshan
Katrina Woodworth


This paper was prepared by SRI International with funds provided by the National Writing Project under a grant from the U.S. Department of Education. However, the contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal Government.
Abstract

By 2013, all 50 states and the District of Columbia had adopted college- and career-ready standards in English language arts and mathematics, placing a greater emphasis on argument writing to prepare students for life after high school. Solving the specific problem of how to help teachers teach to new standards for argument writing as well as the broader problem of improving teaching and learning requires continued efforts to understand how to support teachers in making substantial changes in instruction. This paper reports on a district-randomized controlled trial of the National Writing Project’s College-Ready Writers Program (CRWP), implemented in high-need rural districts in 10 states, testing one instance of professional development paired with supporting curricular resources and a formative assessment tool designed to shift instruction to align with the new college- and career-ready standards in English language arts classes in grades 7–10. Researchers randomly assigned 44 rural districts either to receive 2 years of CRWP or to be in a business-as-usual control group and found positive impacts on teacher practice and student source-based argument writing.
Across the United States, a consensus has formed about the need to better prepare students for academic, career, and civic responsibilities after high school, and new “college- and career-ready” academic standards, including the Common Core standards, have been widely adopted. By 2013, all 50 states and the District of Columbia had adopted college- and career-ready standards in English language arts and mathematics. Even as some states have backed away from standards, assessments, and materials explicitly labeled “Common Core,” the replacements reflect a similar vision for the knowledge and skills students should attain to be successful in the 21st century. One key change in the new standards is the recognition that argument writing needs greater emphasis in curriculum and instruction to prepare students for life after high school.

High-quality professional development is widely seen as critical for supporting teachers in learning the new skills to teach argument writing. Yet substantial research indicates that professional development programs have an inconsistent track record in changing teacher practice and improving student achievement (Desimone & Garet, 2015; The New Teacher Project, 2015). Solving the specific problem of how to help teachers teach to new standards for argument writing as well as the broader problem of improving teaching and learning requires continued efforts to understand how to support teachers in making substantial changes in instruction.

This paper reports on a district-randomized controlled trial of the National Writing Project’s College-Ready Writers Program (CRWP) testing one instance of professional development designed to shift instruction to align with college- and career-ready standards.
CRWP provided 7th- through 10th-grade English language arts (ELA) teachers with 90 hours of professional development over 2 years, curriculum resources for teaching argument writing, and a formative assessment tool focused on techniques for using nonfiction sources to develop an argument. To evaluate the impacts of CRWP, researchers randomly assigned 44 rural districts either to receive 2 years of CRWP or to be in a business-as-usual control group. The study found CRWP to have a positive impact on argument writing instruction and the quality of student argument writing from nonfiction sources. The study tested an approach to professional development that is gathering evidence of success across multiple studies.

**Context for the Study**

We situate this study in the policy landscape, the research base on effective professional development, and the capacity of the program developer—the National Writing Project.

**The policy landscape concerning argument writing**

Writing is an essential skill for participating in modern American society. Although many types of writing are important, argument writing is pervasive in academia, the workplace, and civic society (Graff & Birkenstein, 2010; Smith, Wilhelm, & Fredricksen, 2012). In explaining the “special place of argument writing” in our society, the authors of the Common Core standards first made the case that there is a “unique importance of argument in college and careers” because of how frequently students and those in professions are called on to take facts, make a decision based on those facts, and then explain the rationale for the decision to others (National Governors Association 2010, pp. 24–25). They noted the centrality of argument to research and knowledge production. Additionally, they went on to write:

The value of effective argument extends well beyond the classroom or workplace, however. As Richard Fulkerson (1996) puts it in *Teaching the Argument in Writing*, the
proper context for thinking about argument is one “in which the goal is not victory but a good decision, one in which all arguers are at risk of needing to alter their views, one in which a participant takes seriously and fairly the views different from his or her own” (pp. 16–17). Such capacities are broadly important for the literate, educated person living in the diverse, information-rich environment of the twenty-first century.

Despite this crucial importance, student writing achievement falls far short of national expectations (College Board, 2004; Persky, Daane, & Jin, 2003). For example, on the most recent (2011) NAEP writing assessment, only 27% of 8th-graders and 27% of 11th-graders scored proficient or higher (National Center for Education Statistics [NCES], 2012).

New college- and career-ready standards set ambitious goals for instruction in hopes of improving these outcomes. In particular, the Common Core standards call for teachers to make three main shifts in practice:

- “Regular practice with complex texts and their academic language…”
- “Reading, writing, and speaking grounded in evidence from texts, both literary and informational…”
- “Building knowledge through content-rich nonfiction…. ”

Moreover, the authors of the Common Core standards explained that the “focus on evidence-based writing along with the ability to inform and persuade is a significant shift from current practice” (emphasis added). For teachers to make these significant shifts in their practice, they will need a range of professional supports.

**Prior research on changing instruction**

Professional development is a logical approach to supporting teachers in changing their instruction. The general understanding of “effective” professional development stems from a
survey of a nationally representative sample of more than 1,000 teachers who participated in mathematics and science professional development. That study identified five features thought to make professional development more likely to be effective: content focus, active learning for teachers, coherence with instructional context, sustained duration, and collective participation (Garet, Porter, Desimone, Birman, & Yoon, 2001). Beneath the consensus view that emerged from this study about the features of effective professional development was a much more nuanced story, especially when evidence solely from randomized controlled trials (RCTs) is considered.

In a broad analysis of the research on teacher professional development, Yoon et al. (2007) found nine studies of professional development that met What Works Clearinghouse guidelines. These studies were all relatively small and featured professional development delivered by the study author or affiliated researchers. Although a trend of positive impacts was found, all the studies focused on elementary grades, leaving a complete void in experimental evidence supporting effective professional development for secondary teachers. Subsequently, two large-scale professional development experiments funded by the Institute of Education Sciences (one on middle school mathematics and one on early reading instruction) found no impacts on student achievement even though the programs were carefully designed to include the features of “effective” professional development and were generally implemented as intended (Garet et al., 2008; Garet et al., 2011). These findings, along with more descriptive reports (e.g., The New Teacher Project, 2015) and other more recent RCTs (e.g., Bos et al., 2012; Heller, Daehler, Wong, Shinohara, & Miratrix, 2012; Gallagher, Woodworth, Park, & McCaffrey, 2014) have raised doubts about the efficacy of professional development in supporting teachers in changing instruction.
Yet other recent RCT studies suggest professional development can have measurable effects on student learning when the programs combine professional development with another element—typically curricular materials. These RCTs, all of which found positive impacts on student learning, covered a range of subject areas including civics (Barr et al., 2015), literacy (August et al., 2014), STEM (Newman et al., 2012), economics (Finkelstein, Hanson, Huang, Hirschman, & Huang 2010), writing (Kim et al., 2011), science (Heller et al., 2012; Penuel, Gallagher, & Moorthy, 2011; Taylor, Roth, Wilson, Stuhlsatz, & Tipton 2016), and reading (Gersten, Dimino, Jayanthi, Kim, & Santoro, 2010; Simmons et al, 2010). The programs ranged from 18 hours to 10 days of professional development and included elementary and secondary teachers and students. The RCTs all involved multisite trials.

Collectively, these studies suggest that professional development can impact student learning if teachers acquire new content knowledge, are able to envision what the new practices might look like with their students, and have supports for making sense of new ideas about instruction and transferring these ideas into their classrooms. The results are also aligned with Cohen and Ball’s conception of an instructional triangle composed of interactions between and among teachers, students, and content. Professional development programs that are integrated with supporting instructional materials can target two parts of the triangle—teachers and content—thus increasing the chances for impacting instruction (Cohen, Raudenbush, & Ball, 2003). The study described in this paper tested that hypothesis with a specific National Writing Project program.

**National Writing Project**

The National Writing Project—a network of nearly 185 university-based local Writing Project sites—has been supporting teacher professional development since its origins as the Bay
Area Writing Project in 1974. Local Writing Project sites share a model of university faculty working in collaboration with K–12 expert teachers. The core principle guiding this collaboration is that expert teachers are the best teachers of other teachers. Local Writing Project sites work to develop motivated expert teachers into teacher leaders who can in turn plan and provide customized professional development for other teachers.

The National Writing Project’s national office serves as a hub to spread new and promising ideas throughout the network. Local sites determine which ideas to take up and how to adapt them to their local contexts. Thus, at both the teacher level (teachers within local Writing Project sites) and the site level (local Writing Project sites within the National Writing Project network), the National Writing Project generates and shares ideas using “adaptive scaling” (Mehan, Hubbard, & Datnow, 2010; Rowan & Miller, 2007). Adaptive scaling recognizes the inevitably and embraces the desirability of local adaptation to context. It embodies the idea that the best outcomes can be obtained through adherence to key program principles and adaptation of specific program features to the constraints and affordances of local contexts.

Prior experimental research on National Writing Project professional development programs fit the trend in studies of the impact of professional development on teacher instruction and student learning. Two multisite RCTs of professional development-only programs did not find positive impacts on student outcomes (Gallagher et al., 2012; Gallagher et al., 2014). Another set of RCTs of a local Writing Project site-developed professional development program that includes curricular resources found positive impacts on student learning that have been replicated across multiple studies (Kim et al., 2011; Olson et al., 2012; Olson, et al., 2016). For the current study of the College-Ready Writers Program, funded by an i3 (Investing in Innovation) validation grant, the National Writing Project built on lessons learned from these
earlier studies. The CRWP design featured both central supports for successful adaptive scaling and resources to support participating teachers in making substantial changes to instruction that would impact student writing outcomes within 2 years.

**The College-Ready Writers Program**

The National Writing Project established a leadership team to develop and support CRWP implementation. The leadership team comprised National Writing Project staff and carefully selected teacher leaders and site directors from local Writing Project sites not participating in the study. This team provided the intellectual leadership for the program. Described here is the program they designed and the outcomes they sought to influence—namely, teachers’ instructional practices and students’ proficiency with argument writing (Figure 1). We begin with an overview of the technical assistance the National Writing Project’s CRWP leadership team provided to local Writing Project sites that in turn implemented the program with partner districts in their local service area.

**Technical assistance**

The CRWP leadership team made strategic decisions about how best to support implementation. The team designed structures to support shared learning, developed resources for teacher learning and the transfer of new ideas into the classroom, and established systems for monitoring progress.

The CRWP leadership team designed five national convenings (two at the outset to launch the program, two during the first year of the program, and the final one during the second program year) for all participating local Writing Project sites and district representatives to collaborate on specific aspects of program work. In addition, site leaders attended 1-day convenings connected to the National Writing Project’s Annual Meeting. These convenings
provided information to deepen site and district leaders’ content and pedagogical knowledge about argument writing as well as processes and protocols to facilitate needs assessment and coplanning of the professional development in each district.

Members of the leadership team also assumed roles as “thinking partners”—experienced site leaders deeply versed in the program who had one-on-one phone calls every month with leaders at each local Writing Project site and who visited each site at least annually. The phone calls, visits, and observations at national convenings provided the CRWP leadership team with data on implementation successes and challenges, which they used to determine how best to support sites’ CRWP work.

During Year 1, the primary resource developed for the local Writing Projects was the Online Learning Experience, an online mini-course on teaching source-based argument writing. Each site sent key “teacher consultants” (teacher leaders connected to the local Writing Project sites) who were providing professional development in CRWP districts to the class, ensuring that the local site teams developed a shared knowledge base and language about teaching source-based argument writing.

Over the course of Year 1, the CRWP leadership team carefully monitored program implementation through site visits and regular check-in calls, examined early indicators of progress, and requested formative feedback from its external evaluators. Through this process, the CRWP leadership team realized that teachers would need additional and focused support to ensure meaningful transfer to the classroom within the 2 years allotted for the program.

The leadership team therefore decided to concentrate the Year 2 program exclusively on argument writing and to design curriculum resources to enable teachers to translate ideas they were learning in the professional development into their classrooms. These mini-units, developed
by the current and former teachers on the CRWP leadership team, systematically addressed key ideas in effective argument writing and became a focus of CRWP work in Year 2. Unlike the Online Learning Experience in the first year, these resources directly reached both local Writing Project site teams and teachers in CRWP districts. The leadership team also developed the Using Sources Tool, a formative assessment that guided teachers through a series of prompts to analyze student writing and determine appropriate next instructional steps. Both resources are described below.
Figure 1. College Ready Writers’ Program Logic Model

Context for Teaching and Learning

- Policy context (college- and career-ready standards, aligned assessments and accountability systems)
- Curriculum (existing materials and programs)
- Teacher turnover

Local Writing Project: Key Program Components

Intensive professional development to support classroom implementation
- Focus on college- and career-ready standards-aligned source-based argument writing
- 45 hours per year
- 80% of grade 7–10 ELA teachers participate
- Support for classroom implementation

Curricular resources (Y2)
- Multiday instructional sequences
  - Nonfiction text set
  - Close reading and exploratory writing
  - Argumentation focus
  - Writing processes
  - Provide student work for formative assessment

Formative assessment tools to inform instruction (Y2)
- Analysis of student strengths and areas for improvement
- Focused on student use of source materials
- Used to inform both Writing Project and teachers’ work to meet student needs

NWP Technical Assistance

- Leadership team
- Thinking partners
- National meetings
- Online learning experiences (Y1)
- Curricular resources (Y2)
- Formative assessment tools (Y2)

Teacher Instructional Practices

Teaching argument writing from nonfiction sources
- Close reading and annotation
- Claim and evidence
- Prewriting
- Effective use of source material

Student Outcomes

Performance on on-demand source-based argument writing task

- Content: Writing presents an argument supported by reasoning and developed through the use of evidence from the sources
- Structure: Writing establishes an order and arrangement to enhance the central argument
- Stance: Writing establishes credibility; tone and style are appropriate for purpose
- Conventions: Writing demonstrates age-appropriate control of conventions
Key program components

Year 1. The National Writing Project initially designed CRWP based on widely agreed on characteristics of high-quality professional development—namely, that it be intensive (90 hours over 2 years) and involve collective participation (at least 80% of ELA teachers), focus on specific content (argument and informational writing), and use strategies designed to support classroom implementation (e.g., demonstration lessons, coaching, designing tasks/assignments, and analyzing student work). While committing to these key program components, local Writing Projects had substantial flexibility in designing programs based on their own expertise and their sense of their district partners’ needs.

Year 2. In the summer before Year 2, the CRWP leadership team rolled out supplemental resources—the curricular mini-units and Using Sources formative assessment tool, as well as guidelines for sequencing and using them—to provide additional support for classroom implementation. In addition, the CRWP leadership team decided to narrow the focus of the professional development to argument writing. As a result, by Year 2 CRWP came to be defined by the following interdependent program components:

- Collective participation in intensive professional development to support classroom implementation. The CRWP model continued to call for at least 80% of 7th- through 10th-grade ELA teachers in each district to participate in at least 90 hours of CRWP professional development over 2 years (45 hours per year) and continued to emphasize support for classroom implementation via coteaching, demonstration lessons, coaching, and coplanning with CRWP curricular resources and based on evidence from the formative assessment tools. Local Writing Project site leaders and teacher consultants provided the professional development.
**Curricular resources.** To further support classroom implementation, local Writing Projects were introduced to the set of curricular resources for argument writing. Designed by members of the CRWP leadership team, the CRWP mini-units are multiday instructional sequences that engage students in reading multiple nonfiction texts (a “text set” on a topic of interest to teens), include embedded strategies for reading and comprehending the texts, emphasize one or more key argument writing skills, and call for students to draft and revise their own multi-paragraph argument. Additionally, by addressing particular skills necessary for developing strong written arguments, the mini-units were constructed with the idea that they would be “educative” for participating teachers, whose own content knowledge and pedagogical skills would grow through using them to teach.\(^1\) Local Writing Project sites were expected to spend at least half their professional development events supporting teachers in implementing these curricular resources and to ask teachers to implement four “replacement” mini-units over the course of the year. These mini-units were designed to catalyze the transfer of abstract ideas and new content knowledge into teacher instruction, while the professional development provided critical scaffolds for rapid implementation.

**Formative assessment to inform instruction.** In Year 2, CRWP also incorporated the regular use of the formative assessment tool to focus teachers’ analysis of student work. The NWP designed the Using Sources Tool for teachers to use collaboratively to identify what their students could already do with source-based argument writing and where they need additional teaching. Local Writing Project sites were asked to introduce their

---

\(^1\) For a further discussion of educative materials and how they can be structured to simultaneously build teachers’ and students’ knowledge and skills see, for example, Schneider and Krajcik, (2002).
teachers to the Using Sources Tool in professional development and to support them in using it to look at student work and submit ratings via the National Writing Project’s online portal at least twice before the end of February. The formative assessment tool, in turn, provided teachers with both rapid feedback on student learning as they were implementing new ideas and suggested next steps in instruction—thus feeding critical data to reinforce motivation for uptake and information to guide implementation. The formative assessment tool was also educative, serving to build teachers’ content knowledge by helping them identify qualities of effective arguments in their students’ writing (e.g., the writing “moves” Joseph Harris describes in his 2006 book *Rewriting: How to Do Things with Texts*). Finally, the requirement to submit ratings of student work created a sense of responsibility to ensure use of the Using Sources Tool and thus allow for aggregated results to be shared at national convenings.

The NWP designed these three components of CRWP to reinforce each other to support the development of teachers’ content knowledge and scaffold the transfer of new instructional ideas into teachers’ instructional practice.

**Research Design**

The evaluation was a district-randomized controlled trial measuring the impact of CRWP on districts’ 7th- through 10th-grade ELA instructional practices and source-based argument writing achievement. Given the size of the rural districts (between 1 and 24 ELA teachers in grades 7–10), randomization of entire districts allowed for the development of a community of practice among teachers. The longitudinal aspect of the research design also benefited from randomization at the district level, allowing tracking of students during a 2-year intervention spanning middle and high school (that is, students changed teachers, grades, and even schools
while maintaining their treatment condition). The research team randomized districts within matched pairs to provide better equivalence across baseline indicators of key outcomes and of local context factors. Additionally, this randomization within pairs aided in recruitment and program implementation by ensuring a balanced number of treatment and control districts within local Writing Project sites.

**Recruitment, randomization, and the counterfactual condition**

The National Writing Project recruited 12 local Writing Project sites with a history of leading intensive inservice work and with experience providing professional development on college- and career-ready standards. The local Writing Project sites were in 10 states: Alabama, Arizona, Arkansas, Louisiana, Mississippi, Missouri, New York, Oklahoma, South Carolina, and Tennessee. The National Writing Project instructed local Writing Project sites to recruit pairs of high-need rural districts\(^2\) that were of a similar size, serving comparable proportions of students, and with similar writing outcomes (or ELA outcomes when states did not measure writing). Districts where sites had an established ongoing relationship were not eligible for the study. Each local Writing Project site recruited one to three pairs of districts.

Local Writing Project sites recruited a total of 44 districts (22 pairs). The research team randomized one district in each pair into treatment and control. Treatment and control districts were of similar sizes and served similar percentages of students eligible for free or reduced-price meals, students of color, and English learners (see Table 1).

\(\text{---}\)

\(^{2}\text{Rural districts were defined as eligible for rural federal funding through Small, Rural School Achievement or Rural and Low-Income School programs. The target for differences among districts within each pair was 10 percentage points on quantitative indicators.}\)
Table 1. District Characteristics

<table>
<thead>
<tr>
<th>District-Level Mean Characteristics of CRWP Districts at Baseline (by Treatment Group)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
</tr>
<tr>
<td>Number of students</td>
</tr>
<tr>
<td>Free or reduced-priced meals (%)</td>
</tr>
<tr>
<td>Students of color (%)</td>
</tr>
<tr>
<td>English learners (%)</td>
</tr>
</tbody>
</table>


Control districts were asked to refrain from seeking additional writing professional development aligned with college- and career-ready standards for writing, other than as required by their state, and they were offered a 1-year version of CRWP in the year after the study (business-as-usual with delayed treatment control condition). They received a stipend of $2,500 per semester ($12,500 total from spring of the baseline year through 2 years of treatment), which could be used for other educational purposes for teachers in grades 7–10 (e.g., textbooks, college- and career-ready standards-aligned math professional development). After the 2 years of the evaluation, local Writing Project sites provided a streamlined, 1-year version of CRWP to the control districts. No districts left the study after randomization.

**Teacher sample**

The intervention sample teacher population included all ELA teachers in grades 7–10 who taught a core ELA class (e.g., grade-level English class, excluding electives and support classes) for at least half a year during either study year. The study excluded teachers in schools serving exclusively nontraditional students (e.g., alternative and continuation schools) or those

---

3 The goal of the business-as-usual counterfactual was to simulate what district teachers would have experienced if the National Writing Project had not developed CRWP. The study data showed that typically this meant relatively little professional development on writing instruction and few resources specifically on argument writing. An exception was districts in New York State, where federal Race to the Top funding was used to develop the curricular resource EngageNY, which provided texts and associated instructional materials aligned with the Common Core State Standards. Control districts in the study were free to (and did) use the EngageNY resources.
without ELA departments (i.e., shared-time vocational schools). CRWP called for Local Writing Project sites to serve all target grade 7–10 ELA teachers in CRWP districts and permitted sites to offer professional development to any teachers in CRWP districts outside the target population (e.g., social studies and elementary school teachers).

**Student sample**

To align with the study goal of measuring CRWP’s impact on writing achievement in study districts after 2 years, researchers defined the student study population as 7th- to 9th-grade students in eligible schools within each district in Year 1 (fall 2013) who took the baseline writing assessment (described below). (Teachers were directed to exclude students exempt from state testing in English, as the assessment would be inappropriate for them.)

Given the expense of scoring student writing, we could not score all papers from the study population of students; power estimates indicated that roughly 60 students per district would provide a minimum detectable effect size of .20 or below. To select an unbiased sample to estimate student writing in each district, we assigned a random number to all students for whom we collected a baseline prompt in fall 2013. We used this preassigned random number to sample 60 students per district (as available) who had complete pre- and post-test writing. Only writing from these sampled students was scored.4

---

4 In the five districts with fewer than 60 students, we scored writing from all students who completed both a pre- and post-test. In the other 39 districts, we scored writing from 60 students per district, stratifying samples within districts and using the preassigned random number to select students. The research team stratified the student sample into cells for sampling by (1) district, (2) baseline grade, (3) baseline prompt type (two were given in each grade—one argument, one informational), and (4) outcome prompt type (two were given in each grade—both argument). This stratification created 12 cells per district. Researchers ranked all students within their individual cells by their pre-assigned random number. Given CRWP’s focus on argument writing in Year 2, we first sampled students from the six cells including a baseline measurement of argument writing to provide a better aligned measure of baseline student achievement. The first 10 students from each of these six cells (ranks 1-10) were automatically included in the student sample. For every student in these first 10 without a Year 2 prompt, we sampled the next ranked student until reaching 10 students with both baseline and outcome data. If these argument cells contained fewer than 10 students with both a pre- and post-test, we sampled from the equivalent (by district, grade, and outcome type) cell for students who wrote a baseline informational prompt.
Inclusion of in-moving teachers and students

The analytic samples of teachers and students both contained individuals who moved into the district after the announcement of randomization decisions (often referred to as in-movers or joiners). Randomization decisions were announced to districts in the spring of 2013 to enable districts and local Writing Project sites to begin coplanning the teacher professional development. The program began in summer 2013, and student baseline data (used to define the analytic sample of students) were not collected until fall 2013. The study included in-moving teachers for the sake of collecting data that accurately represented the instructional practices in each district throughout the study. If we had excluded in-mover teachers, our data would not have provided a representative description of teaching practice within the districts. In-movers could have been a threat to randomization if students or teachers changed districts to take advantage of the treatment being offered.5

Data and methods

Over the 2-year study, researchers collected data to document (1) CRWP program implementation (professional development monitoring and interviews and observations), (2) the difference in the amount of and nature of the professional development experienced by CRWP and control teachers (teacher survey), (3) changes in teachers’ classroom instruction (teacher survey and log), and (4) students’ writing performance (on-demand writing prompts). This paper

5 As such, these results may contain both the effects of assignment to CRWP on measured outcomes for teachers and students and any effects this assignment had on the composition of the teaching force and student body unmeasured and unaccounted for by baseline equivalence measures. That is, teachers and parents may have moved into treatment districts to benefit from CRWP. The particular circumstances of this project minimize the potential for bias by in-moving students. We randomized at the district level in rural communities—so parents’ choices to live in one district over another are unlikely to be influenced by the treatment, substantially reducing the possibility of bias on student outcomes.
reports on Year 2 findings, which represent the cumulative impact of the program on teachers and students. In this section, we describe the data collection activities on which we report.

**Professional development monitoring**

To document program implementation, local Writing Project leaders completed a professional development monitoring form to capture information about each professional development event (the format, content, and duration) and to record teacher participation. Site leaders submitted the forms three times per year (at the end of summer, fall semester, and spring semester).

**Interviews and observations**

To gather more nuanced data on program implementation, researchers interviewed teachers, school and district administrators, and local Writing Project leaders and observed classrooms and CRWP professional development. Researchers visited the CRWP districts once or twice per year and periodically conducted phone interviews with key informants (e.g., local Writing Project site directors and school leaders). To deepen understanding of treatment-control contrast, researchers also visited control districts to conduct interviews and classroom observations, albeit less frequently and with smaller samples. In this paper, we only briefly draw on this qualitative data.

**Teacher log and survey**

Teacher measures were obtained from a daily instructional log and survey adapted from instruments with strong reliability from a prior study (Cronbach’s α’s between .83 and .96; Gallagher et al., 2012), with modifications for better alignment with the teacher instructional
practices targeted by CRWP.\(^6\) Researchers administered the online instructional log once a day over 3 weeks, with a baseline week in spring 2013 and outcome weeks in fall and spring of the 2014–15 school year. Researchers administered the online teacher survey at baseline in spring 2013, before the beginning of the program, and again as an outcome measure in spring of Year 2. All administrations were timed to avoid extended school breaks and standardized testing windows.

On the daily instructional log, teachers recorded time spent writing, length of writing assigned, and the purposes for writing during instructional time that day. Researchers asked both CRWP and control teachers to fill out the log daily to more accurately capture the specific classroom practices enacted and sampled multiple days to minimize the error associated with the measurement of teacher practice. To improve the precision of the estimates, the log prompted teachers to focus on an average individual student in a particular class rather than estimate across students. In addition, since most teachers taught multiple classes eligible for study, we sampled different classes on different days of the week.

The survey measured broader practices and constructs more appropriately measured over a year than in a single day. Both CRWP and control teachers were asked identical questions about the most intensive writing activity they asked of their students that year, relative emphasis on facets and genres of writing, perceptions of self-efficacy and the policy environment, and experiences in professional development. CRWP teachers also received a series of questions

\(^6\)To assess the interrater reliability of the daily log used in this study, pairs of researchers observed classrooms in spring of Year 2. The researchers and the teacher agreed on a focal student at the beginning of the class. Each researcher filled out one log, as did the teacher being observed. Researchers observed 36 classes across CRWP and control districts. Researchers compared results from these three logs to assess how frequently they agreed with each other about the activities observed and the extent to which teachers agreed with researchers about class activities for the day. Teachers and researchers all demonstrated high interrater reliability on each of the questions, with all interrater reliabilities greater than 65% and the majority more than 80%.
about specific program elements (e.g., use of the mini-units and Using Sources Tool) to better understand program uptake among these teachers.

**Student writing data**

To measure students’ writing ability, researchers administered on-demand assessments of source-based argument or informational writing to all students in grades 7–9 as a baseline measure in fall of Year 1. In spring of Year 2, researchers administered an on-demand assessment of source-based argument writing to all students in grades 8–10 as an outcome measure. The assessments provided students with four to six short texts to read, and the outcome prompt asked students to write an argument using evidence from the texts. The student writing was scored with the Analytic Writing Continuum for Source-based Argument.

Over a decade, the National Writing Project developed the Analytic Writing Continuum, which has been shown to be a valid and reliable measure of student writing (Bang, 2013). The original version of the Analytic Writing Continuum had been used primarily to score writing rooted in students’ personal experience and therefore did not explicitly measure the use of evidence from other sources. For the CRWP evaluation, the National Writing Project developed writing prompts that would require students to select and use evidence from written sources to support their own claims or to inform an audience about a particular issue. The resulting performance tasks are similar to performance-based tasks that were part of some state assessments (e.g., Connecticut) and are part of national assessment consortia (i.e., PARCC and

---

7 State standardized tests, when available and aligned with the intervention, provide an efficient and policy-relevant student outcome for teacher professional development studies. However, when this study was conducted, most participating states’ ELA tests did not require students to engage in extended argument writing, much less on an annual basis as would be needed to provide a suitable outcome measure for the study.

8 For more information about the National Writing Project Analytic Writing Continuum, see [http://www.nwp.org/cs/public/print/resource/3776](http://www.nwp.org/cs/public/print/resource/3776)
The National Writing Project worked with a panel of writing assessment experts to modify the Analytic Writing Continuum to more accurately score writing that relied on the use of external sources as evidence. The same panel of writing assessment experts selected and annotated anchor papers to be used in training scorers. The revisions to the Analytic Writing Continuum and the development of annotated anchor papers were designed to help make explicit for scorers how well-established attributes of effective writing are evident in source-based argument writing. For example, the AWC-SBA’s rubric for the stance attribute directs reviewers to assess the extent to which the writing establishes the credibility of the cited source material. The resulting Analytic Writing Continuum for Source-based Argument (AWC-SBA) retains a basic structure rooted in the “six traits” of writing but has a particular focus on the attributes related to source-based argument writing. The AWC-SBA measures four attributes: content (e.g., quality of reasoning and strength of evidence), structure (e.g., organization to enhance the argument), stance (e.g., tone, establishment of credibility), and conventions (e.g., control of usage, punctuation, spelling, capitalization, and paragraphing).

For unbiased administration and scoring, local research site coordinators were hired to support and monitor data collection in person in all schools and return the completed prompts to the research team. Researchers sampled student papers for scoring (described above) and then de-identified the samples by removing names or other personal identifying information and assigning an anonymized identification number. The de-identified papers were sent to the National Writing Project for scoring. Scorers therefore did not know the district the papers came

---

9 One local Writing Project site sent staff to treatment districts during administration of the outcome prompt. To verify that this presence did not inadvertently bias the outcome findings, researchers estimated a model with these districts dropped from the analysis sample. The impacts on content and structure were unaffected by the exclusion of the districts; impacts on stance and conventions were qualitatively similar, although the impact on stance was only marginally significant (p < .1).
from or the treatment status. Scorers were recruited from current and former teachers affiliated with local Writing Project sites not participating in CRWP (to limit the potential for bias in scorers familiar with the program). Researchers monitored the scoring to ensure the National Writing Project followed impartial processes.

Table 2. Students’ Scores on Analytic-Writing Continuum for Source-based Writing: Means by Treatment Condition and Time

<table>
<thead>
<tr>
<th></th>
<th>Baseline, Unstandardized</th>
<th>Outcome, Unstandardized</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment</td>
<td>Control</td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>SD</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>n</td>
<td>1,259</td>
<td>1,227</td>
</tr>
<tr>
<td>Structure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>6.0</td>
<td>5.5</td>
</tr>
<tr>
<td>SD</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>n</td>
<td>1,259</td>
<td>1,227</td>
</tr>
<tr>
<td>Stance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.7</td>
<td>2.6</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>SD</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>n</td>
<td>1,259</td>
<td>1,227</td>
</tr>
<tr>
<td>Conventions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>2.7</td>
<td>2.8</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>6.0</td>
<td>6.0</td>
</tr>
<tr>
<td>SD</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>n</td>
<td>1,259</td>
<td>1,227</td>
</tr>
</tbody>
</table>

NOTE: Mean scores do not adjust for grade or prompt effects.

Reliability of the prompt scoring was assessed separately for each writing attribute measure in the AWC-SBA through the double scoring of a subset of papers. Researchers
randomly selected approximately 15\% of the papers to be double scored and calculated the percentage of papers for which individual scorers agreed within a score point for each attribute. These interrater reliabilities therefore provide estimates of the reliability of the 85\% of writing not double scored. A total of 702 papers in the analytic sample were double scored; raters agreed within a single score point for 91\% of papers on the content and conventions attributes, 92\% on the structure attribute, and 89\% on the stance attribute.

**Models**

The predicted writing ability for student $i$ in district $j$ in blocking pair $k$ as a function of attending a district assigned to treatment is given as

$$Y_{ijk} = \beta_0 + \beta_1(Treatment_j) + \beta_2(Pre - Test_i) + \epsilon_{ijk} + \eta_{jk} + \mu_k.$$  

$Y_{ijk}$, student $i$’s writing ability at the end of Year 2, was standardized within the outcome prompt and cohort grade to provide an effect size that accounted for the distribution of mean scores within each outcome prompt taken. Random effects $\epsilon_{ijk}$, $\eta_{jk}$, and $\mu_k$ allowed for error at the student, district, and block level, respectively. The main impact model included a control for student baseline achievement in fall of Year 1. These student baseline scores were standardized within cohort (i.e., 7th–9th grade at baseline) and baseline prompt to account for prior achievement, cohort at baseline, and prompt taken at baseline. $\beta_1$ provided an estimate of the effect of district assignment to treatment on student writing ability within that district (the intent-to-treat effect).

We estimated these multilevel models using the Stata 14.1 *mixed* command. They used restricted maximum likelihood estimation and the Kenward-Roger method to compute degrees of freedom for the models and calculate $p$-values to adjust for the relatively small sample size at the
Student achievement was similar across CRWP and control districts at baseline (see Table 2). Researchers further established baseline equivalence of student writing using the same structural model but predicting the student pretest score; all $p$-values for these models exceeded .1. These estimates demonstrate that student achievement was statistically similar in each of the districts at baseline, adjusting for the clustering of students within districts.

To estimate the impact of CRWP on teacher outcomes, we compared survey and log indicators of teaching practice using models similar to those used to analyze student outcomes. Survey data also required a three-level hierarchical linear model (blocking pair, district, and teacher), whereas the logs required a fourth level in the models (the logs themselves), given the multiple measurements per teacher. To compare binary outcomes, we used `melogit` in Stata version 14.1. For ease of interpreting outcomes with logistic mixed-effect models, we translated model-predicted mean logits to predicted probabilities for the average teacher or day in each condition.

**Results**

In presenting study results, we start with program implementation and then move to the treatment-control contrast in professional development experiences and instruction and the resulting impacts on student writing.

**CRWP Implementation**

Despite the challenge of implementing a consistent program in 22 districts across 10 states delivered by 12 Writing Project sites, local Writing Projects maintained high fidelity to
key CRWP components. Here, we describe fidelity of implementation as measured using key indicators for each of the three program components.

**Year 1.** In examining fidelity of implementation at the end of Year 1, we found that large majorities of teachers in nearly every district experienced a program defined by key program components:

- In 20 of 22 districts, 80% or more of ELA teachers participated in 45 or more hours of CRWP professional development.
- In 16 of 22 districts, 80% or more of ELA teachers participated in 18 or more hours of CRWP professional development focused on argument or informational writing.
- In 22 of 22 districts, 80% or more ELA teachers participated in 8 or more hours of professional development designed to support classroom implementation.

**Year 2.** Because of the substantial focusing of the program in Year 2, we provide more detail on the fidelity of implementation of the Year 2 program components: intensive professional development and use of curricular resources and a formative assessment tool.

- **Intensive professional development designed to support classroom implementation.** As in Year 1, across the CRWP districts 20 of 22 districts met the implementation fidelity threshold for participation (80% of ELA teachers participating in 45 or more hours). Just half the districts (11 of 22) reached the threshold for participation over both years (90 hours or more over 2 years) because of instability in the teacher workforce; because of turnover, many teachers were not present in the schools long enough to experience the full 2-year intervention.

  In Year 2, the CRWP leadership team expected half the CRWP professional development events teachers experienced to be delivered using approaches to professional
development that would support classroom implementation of argument writing—in particular coaching, coteaching, and demonstration lessons; coplanning and codesigning learning tasks and assignments; and analyzing student work. Eighteen of the 22 districts met this threshold (80% of ELA teachers experiencing at least half their CRWP professional development in the format of one of these formats). Moreover, nearly three-fourths (73%) of teachers reported (on the Year 2 teacher survey) observing Writing Project staff model the use of text-based argument tasks and/or benefiting from coaching or coteaching support from the Writing Project as they implemented CRWP text-based argument tasks.

- **Curricular resources.** In Year 2, the CRWP leadership team expected half of CRWP professional development events to specifically support the use of CRWP curricular resources—not only resources developed by the CRWP leadership team, but also those developed by teacher leaders at local Writing Project sites—designed to teach argument writing from nonfiction sources. Eighteen of 22 districts met this threshold; that is, for 80% of district ELA teachers, at least half of their CRWP professional development was on supporting the use of CRWP curricular resources. Further, on the Year 2 survey, 89% of treatment teachers reported that planning with the Writing Project and/or colleagues supported their use of CRWP resources.

- **Formative assessment.** In Year 2, the CRWP leadership team asked local Writing Project sites to introduce ELA teachers to the Using Sources Tool and request that they submit their analysis of student work via the National Writing Project’s online portal on two occasions by the end of February 2015. In 21 of 22 districts, local Writing Projects introduced the Using Sources Tool to at least 80% of the ELA teachers; in all 22 districts, teachers submitted their analysis of student work via the National Writing Project’s online portal on at least two
occasions. In nearly all cases, teachers analyzed student work collectively during a professional development event.

**Differences in treatment and control teachers’ professional development experiences**

There were large, statistically significant differences between teachers in CRWP and control districts in both the quantity and content of the professional development they received (see Table 3). To better understand the contrast between the treatment and control conditions, we asked teachers about their professional development experiences during Year 2. CRWP teachers reported receiving nearly 10 times as many hours of writing professional development as teachers in control districts during Year 2 (63 hours vs. 6.4 hours) and were more likely to report participating in professional development on lessons for teaching argument writing (98% vs. 47%) and on writing from source material (91% vs. 50%). CRWP teachers were also more likely to report participating in professional development that had features consistent with the key components of CRWP such as coteaching or collaborating with an instructional coach and analyzing student work. Moreover, 93% of CRWP teachers reported teaching at least one CRWP mini-unit, and the same percentage reported analyzing student writing with the CRWP Using Sources Tool.
Table 3. Differences in Writing Professional Development Received by Treatment Status, Year 2

<table>
<thead>
<tr>
<th>Since the end of last school year, how many hours of PD on writing instruction have you received?</th>
<th>Treatment</th>
<th>Control</th>
<th>Block n</th>
<th>District n</th>
<th>Teacher n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>63.41</td>
<td>6.41</td>
<td>***</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Received PD that provided lessons...for teaching focused on argument writing (1=Yes)</td>
<td>98%</td>
<td>47%</td>
<td>***</td>
<td>22</td>
<td>42</td>
</tr>
<tr>
<td>Received PD focused on writing from source material (1=Yes)</td>
<td>91%</td>
<td>50%</td>
<td>***</td>
<td>22</td>
<td>42</td>
</tr>
</tbody>
</table>

To what extent did the PD involve...\(^\text{a}\)

|                                                                                              | Treatment | Control | Block n | District n | Teacher n |
| Participating in a lesson, activity, or strategy as a learner                               | 2.82      | 2.21    | ***     | 22         | 42        | 213       |
| Engaging in writing myself                                                                   | 2.65      | 1.83    | ***     | 22         | 42        | 212       |
| Reading and/or discussing professional literature                                            | 2.23      | 1.93    | **      | 22         | 42        | 214       |
| Analyzing student work                                                                       | 2.86      | 2.07    | ***     | 22         | 42        | 212       |
| Planning how to implement what was learned in PD in my classroom                             | 2.78      | 2.50    | ***     | 22         | 42        | 213       |
| Designing tasks/assignments                                                                 | 2.54      | 2.25    | **      | 22         | 42        | 214       |
| Observing a demonstration lesson in a classroom with students                                | 2.11      | 1.60    | ***     | 22         | 42        | 212       |
| Coteaching or collaborating with an instructional coach                                       | 2.26      | 1.47    | ***     | 22         | 42        | 212       |
| Integrating or adapting new approaches to writing instruction with existing materials       | 2.56      | 2.10    | ***     | 22         | 42        | 214       |
| Sharing my expertise with other teachers                                                     | 2.36      | 1.79    | ***     | 22         | 42        | 213       |

To what extent do you dis(agree): I received adequate professional development to teach writing\(^{b}\)

|                                                                                              | Treatment | Control | Block n | District n | Teacher n |
|                                                                                              | 4.27      | 2.97    | ***     | 22         | 44        | 305       |

NOTE: Point estimates for models predicting binary outcomes were transformed into percentage points for ease of interpretation and represent the predicted practice for the average teacher in the sample. Data collected using Year 2 teacher survey.

\(^a\) Scale: 1 = Not at all, 2 = Minor extent, 3 = Major extent

\(^b\) Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree

\(p < .05\). \(**p < .01\). \(***p < .001\).
Impacts on teacher practice in CRWP districts

As a result of CRWP, teachers’ practice changed in focus, with teachers using writing instruction time to practice the elements of argument writing emphasized by the program. Table 4 provides outcomes from the daily instructional log, which demonstrated that the time teachers spent on writing differed very little: Relative to those in control districts, CRWP teachers asked students to write somewhat more frequently (94% of days vs. 89% of days, \( p < .05 \)) and over similar periods of time (about 30 minutes). When students wrote, however, the work they were asked to do differed more dramatically. When asking students to write, CRWP teachers were more likely than teachers in control districts to focus on the facets of argument writing emphasized in CRWP and taught through the mini-units—developing a claim, evaluating evidence that could be used in support of this claim, developing an argument in support of the claim, and practice writing the argument.
Table 4. Differences in Daily Student Writing by Treatment Status, Year 2

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
<th>Block</th>
<th>District</th>
<th>Teacher</th>
<th>Log n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you ask students to write? (1 = Yes)</td>
<td>94%</td>
<td>89%</td>
<td>*</td>
<td>22</td>
<td>44</td>
<td>341</td>
</tr>
<tr>
<td>How long did they write (in minutes)?</td>
<td>32.66</td>
<td>31.11</td>
<td>*</td>
<td>22</td>
<td>44</td>
<td>334</td>
</tr>
<tr>
<td>How much did they write?</td>
<td>2.80</td>
<td>2.50</td>
<td>**</td>
<td>22</td>
<td>44</td>
<td>332</td>
</tr>
</tbody>
</table>

When asked to write, did student work include:

- Practice writing an argument: 26% vs. 7% (***), Block n = 22, District n = 44, Teacher n = 334, Log n = 2,463
- Generate an argument in support of a claim: 26% vs. 10% (***), Block n = 22, District n = 44, Teacher n = 334, Log n = 2,463
- Identify a claim and evidence in text: 27% vs. 26%, Block n = 22, District n = 44, Teacher n = 334, Log n = 2,463
- Develop a claim: 36% vs. 17% (***), Block n = 22, District n = 44, Teacher n = 334, Log n = 2,463
- Evaluate the credibility of evidence: 19% vs. 10% (**), Block n = 22, District n = 44, Teacher n = 334, Log n = 2,463
- Support a claim with evidence from a text: 44% vs. 41%, Block n = 22, District n = 44, Teacher n = 334, Log n = 2,463
- Support a claim with evidence from personal experience: 13% vs. 6% (**), Block n = 22, District n = 44, Teacher n = 334, Log n = 2,463
- Elaborate upon evidence used to support a claim: 24% vs. 15% (**), Block n = 22, District n = 44, Teacher n = 334, Log n = 2,463

NOTE: Logs that answered *students did not write in class* were not included in subsequent analyses; point estimates for models predicting binary outcomes were transformed into percentage points for ease of interpretation and represent the predicted practice for the average teacher in the sample. Data collected using the daily teacher log in Year 2.

a Scale: 1 = One or more single-sentence responses, 2 = Less than a page, 3 = One page, 4 = Two to three pages, 5 = Four to five pages, 6 = More than five pages.

*p < .05. **p < .01. ***p < .001.
The teacher survey revealed that CRWP teachers placed more emphasis throughout the school year on the facets of writing necessary to produce a well-reasoned and well-supported argument than teachers in the control group (Table 5). As in the daily practice logs, we see that CRWP teachers were more likely than control teachers to emphasize the main elements of argument writing developed in CRWP: developing a claim, connecting evidence to a claim, selecting evidence from source material, and introducing and commenting on source material. In addition, 82% of CRWP teachers reported that the most intensive writing assignment they gave students during the year was argument writing, compared with 39% of control teachers ($p < .001$). In addition, CRWP district teachers reported a greater emphasis on on-demand writing in response to text. Treatment teachers may have reported this emphasis as a result of CRWP’s mini-units designed to give students opportunities to practice for on-demand writing situations. While this task could be perceived as “test prep,” the alignment of the outcome task with college- and career-ready standards means that it may prove useful training for situations students will encounter in professional and higher education settings. There is also evidence of commonalities in the writing instruction across CRWP and control district. Teachers in both conditions reported similar emphasis on using language effectively; using words, phrases, and clauses to link the major sections of text; using a style and tone appropriate for the audience and purpose; writing introductions and conclusions; and organizing ideas.

Given that CRWP district teachers reported relatively similar time on task but different focus, treatment teachers would have had to replace some existing practices. From site visit observations, researchers hypothesized that much of the teaching being replaced in ELA teachers’ classrooms would be on conventions and literary analysis. The annual survey findings indicated that CRWP did replace these two facets of work. Control district teachers reported a
greater emphasis on teaching conventions (4.0 vs. 3.6 on a 5-point scale, \( p < .001 \)), and they were nearly four times more likely than CRWP teachers to report that the most intensive writing assignment given to students was in literary analysis (19\% vs. 5\%, \( p < .001 \)). Although these findings are logically consistent with the qualitative findings on site, the sizes of the findings indicate that these substitutions may not completely explain the changes in teacher practice. Qualitative data suggest one possible explanation, namely that teachers integrated nonfiction texts into units on core literature (e.g., interspersing a CRWP mini-unit that included texts on brain development in teens with a unit on *Romeo and Juliet*). Such integration would not necessarily reduce the number of days teachers taught literary analysis of *Romeo and Juliet*, but it would change the nature of that instruction.
Table 5. Differences in Annual Focus of Teachers’ Instruction by Treatment Status, Year 2

<table>
<thead>
<tr>
<th>When writing, how much emphasis did you place on...</th>
<th>Treatment</th>
<th>Control</th>
<th>Block n</th>
<th>District n</th>
<th>Teacher n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing a claim</td>
<td>4.71</td>
<td>4.04</td>
<td>***</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Connecting evidence to a claim</td>
<td>4.76</td>
<td>4.19</td>
<td>***</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Selecting evidence from source material</td>
<td>4.63</td>
<td>4.15</td>
<td>***</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Introducing and commenting on quoted text or source material</td>
<td>4.49</td>
<td>3.68</td>
<td>***</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Using language effectively</td>
<td>3.90</td>
<td>3.96</td>
<td></td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Using words, phrases, and clauses to link the major sections of text</td>
<td>3.66</td>
<td>3.66</td>
<td></td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Using a style and tone appropriate for the audience and purpose</td>
<td>3.86</td>
<td>3.77</td>
<td></td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Writing introductions and conclusions</td>
<td>4.18</td>
<td>4.14</td>
<td></td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Organizing ideas</td>
<td>4.43</td>
<td>4.31</td>
<td></td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Conventions and usage</td>
<td>3.59</td>
<td>3.97</td>
<td>***</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>On-demand writing in response to text</td>
<td>4.04</td>
<td>3.67</td>
<td>**</td>
<td>22</td>
<td>44</td>
</tr>
</tbody>
</table>

Which of the following best characterizes the most intensive writing assignment you gave students this year?b

<table>
<thead>
<tr>
<th>Argument</th>
<th>%</th>
<th>%</th>
<th>Block n</th>
<th>District n</th>
<th>Teacher n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argument</td>
<td>82%</td>
<td>39%</td>
<td>***</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td>Literary Analysis</td>
<td>5%</td>
<td>19%</td>
<td>***</td>
<td>22</td>
<td>44</td>
</tr>
</tbody>
</table>

NOTE: Point estimates for models predicting binary outcomes were transformed into percentage points for ease of interpretation and represent the predicted practice for the average teacher in the sample. Data collected using teacher survey in Year 2.

a Scale: 1 = No or almost no emphasis, 2 = Minor emphasis, 3 = Some emphasis, 4 = Significant emphasis, 5 = Heavy emphasis
b Full list of options included narrative, informational, and other.
* p < .05. ** p < .01. *** p < .001.
Impacts on quality of student writing in CRWP districts

CRWP had a positive, statistically significant, and robust effects on the quality of treatment students’ writing in three of the four attributes measured by the AWC-SBA (Table 6). We estimate an impact of .20 on the content attribute of the AWC-SBA ($p < .05$). The impact estimate on the structure attribute is similar in size and significance. The point estimate for stance is qualitatively smaller, although not statistically different from those estimated on content and structure. The impact estimate on conventions is only marginally significant ($p < .1$). Estimates of impacts on content, structure, and stance are robust to alternate model specifications (Table 7).

Table 6. Differences in Student Writing as Measured by AWC-SBA by Treatment Status, Year 2

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Content</th>
<th>Structure</th>
<th>Stance</th>
<th>Conventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impact estimate</td>
<td>0.20 *</td>
<td>0.20 *</td>
<td>0.15 *</td>
<td>0.12 ~</td>
</tr>
<tr>
<td>SE</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>N's Block</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>District</td>
<td>44</td>
<td>44</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>Student</td>
<td>2,486</td>
<td>2,486</td>
<td>2,486</td>
<td>2,486</td>
</tr>
<tr>
<td>df, Impact estimate</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

NOTE: Outcome scores are standardized within the outcome prompt and cohort grade to provide an effect size that accounted for the distribution of mean scores within each outcome prompt taken. 
~$p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$. 
Table 7. Differences in Student Writing as Measured by AWC-SBA by Treatment Status using Various Model Specifications, Year 2

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact estimate</td>
<td>0.20 *</td>
<td>0.18 *</td>
<td>0.17 *</td>
<td>0.22 *</td>
</tr>
<tr>
<td>SE</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td>(0.08)</td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact estimate</td>
<td>0.20 *</td>
<td>0.19 *</td>
<td>0.18 *</td>
<td>0.20 *</td>
</tr>
<tr>
<td>SE</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.08)</td>
</tr>
<tr>
<td><strong>Stance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact estimate</td>
<td>0.15 *</td>
<td>0.13 *</td>
<td>0.13 *</td>
<td>0.16 ~</td>
</tr>
<tr>
<td>SE</td>
<td>(0.07)</td>
<td>(0.06)</td>
<td>(0.06)</td>
<td>(0.08)</td>
</tr>
<tr>
<td><strong>Conventions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact estimate</td>
<td>0.12 ~</td>
<td>0.13 *</td>
<td>0.13 *</td>
<td>0.11</td>
</tr>
<tr>
<td>SE</td>
<td>(0.07)</td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.08)</td>
</tr>
<tr>
<td><strong>Random block effects</strong></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Fixed block effects</strong></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Student-level prior achievement control</strong></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>District-level prior achievement control</strong></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Outcome scores are standardized within the outcome prompt and cohort grade to provide an effect size that accounted for the distribution of mean scores within each outcome prompt taken.

*p < .1. **p < .05. ***p < .001.
Across multiple attributes, these results provide rigorous evidence of the efficacy of CRWP for improving student source-based argument writing. A recent meta-analysis of effect sizes in relation to methodological features found the effect sizes of evaluations of interventions targeting student achievement in reading, math, or science tended to be smallest in randomized controlled trials and studies with large sample sizes, which the authors defined as $N > 250$ (Cheung & Slavin, 2015). For large sample RCTs, the authors classified .11 as a large effect size—though they noted that the effect sizes for “experimenter-made” assessments tended to be, on average, about twice as large as those of “standardized” measures (e.g., state assessments). Thus, we argue that a .20 effect on this complex and important skill is large enough to be not only statistically significant, but also policy relevant.

**Discussion**

CRWP had large, positive impacts on student outcomes on a particularly complex task—writing an argument supported by reasoning and developed through the use of evidence from source material. This type of argument writing has been identified as critical to college and career readiness and thus is central to new state standards for English language arts and literacy. Moreover, the National Writing Project demonstrated these outcomes at a large scale (22 districts working with 12 Writing Project sites in 10 states) and in challenging settings (rural high-poverty secondary schools).

Of particular note is the issue of teacher turnover, which has been posited as a likely cause of no effects findings in some prior randomized controlled trials of professional development programs (Desimone & Garet, 2015; Garet, et al., 2011; Garet et al., 2008). Teacher turnover within CRWP districts was a challenge to implementation of the 2-year program. Of the 167 target ELA teachers at the start of the professional development, 31% were
no longer teaching target classes in their districts when the study ended in the spring 2015 (e.g., they had left teaching, left their district, switched grades so that they were no longer teaching grades 7–10, or switched subjects).\textsuperscript{10} Turnover is a problem in changing teacher practice and student outcomes for at least two reasons. First, the investment districts and professional development providers make in teacher capacity can be realized only to the extent that teachers remain in the profession (and will extend to the district’s students only if they remain in that district). Second, if teachers move in and out of the subject or grade, their ever-changing needs and context—and the constant fluctuation of the community of practice created—may lessen the efficacy of the professional development provided to them. CRWP led to positive effects despite the fact that roughly one-third of teachers received only 1 year of the 2-year program.

What seems to have made CRWP effective was its three-pronged structure: (1) intensive professional development in support of classroom implementation of new content-specific instructional practices, (2) curricular resources, and (3) formative assessment. The design drew on sound theory about the how to influence instruction that could impact students. In short, the path to improving student outcomes through teacher professional development alone can be slow and recursive as teachers are exposed to new ideas, make sense of them, try them in their classrooms, and reflect—sometimes in collaboration with peers or more expert others—until they successfully bring them into their instructional practices over time. In CRWP, the three interdependent prongs reinforced each other to catalyze this process with (1) formative assessment tools that highlighted key qualities of effective argument writing and helped teachers find evidence for whether their students’ writing had those qualities, (2) curricular resources for

\textsuperscript{10} Desimone and Garet (2015) describe teacher turnover as a challenge facing urban schools. These data also serve as a reminder that it is also a challenge in high-poverty rural districts.
WORKING PAPER: IMPACT OF NWP’S COLLEGE-READY WRITERS PROGRAM

teachers to plan and deliver instruction on key skills, and (3) professional development that supported teachers in transferring new instructional ideas into their classrooms.

Even a well-designed program is difficult to implement at this scale, and the National Writing Project drew on the experience within its network and its orientation toward continuous learning and improvement. The CRWP leadership team members drew on both their content expertise and the results of prior studies of Writing Project programs to design a program that could be implemented with fidelity while maintaining core National Writing Project principles. The National Writing Project took advantage of its broad infrastructure and tapped its own experts who developed an initial vision for the program, monitored local implementation, and rapidly altered the program so that it could be implemented in such a way as to influence teacher practice sufficiently enough to impact student outcomes in a short time. This study of CRWP also provides positive evidence of the National Writing Project’s ability to rapidly scale complex work across multiple sites.

The limitations to these findings are relatively few. These findings can be generalized to a particular population: rural districts, primarily (but not exclusively) in the Southeast. However, the program was enacted consistently in more than 20 districts in 10 states, so we see no reason why it would not be equally successful in other settings. As the study included in-movers, we cannot eliminate the possibility that some teachers and students may have entered the district to benefit from this professional development, but this seems unlikely, especially in the rural settings of these districts where relocation and commuting are more difficult than in most areas.

This study is one of the largest and most rigorous studies of teacher professional development to find evidence of an impact on student academic outcomes. Furthermore, the rich quantitative and qualitative data collected about program implementation and teacher
instructional practice support existing theory about high-quality professional development. This study adds rigorous experimental evidence to a body of literature that indicates that professional development that addresses more than one aspect of instructional capacity—in this case, teacher knowledge and skill, instructional materials, and formative assessment tools—is more likely to lead to meaningful student learning. The findings therefore have important implications for practitioners, policymakers, and program developers.
References


