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

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.

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

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 how to teach argument writing effectively. Yet substantial research indicates that professional development programs have an inconsistent track record in changing instruction 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, SRI Education 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 consider 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, 2012).

New 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…”
- “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 considering evidence solely from randomized controlled trials (RCTs).

In a broad review of the research on teacher professional development, Yoon, Duncan, Lee, Scarloss, and Shapley (2007) found nine studies 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 void in experimental evidence...
around professional development for secondary teachers. Subsequently, three large-scale professional development experiments funded by the Institute of Education Sciences (one on elementary school mathematics, 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; Garet et al., 2016). 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; Gallagher, Woodworth, Park, & McCaffrey, 2014; Heller, Daehler, Wong, Shinohara, & Miratrix, 2012) have raised doubts about the efficacy of professional development in supporting teachers in changing instruction in ways that positively impact student learning.

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 transferring these ideas into their classrooms. The results are also aligned with Cohen, Raudenbush, and Ball’s (2003) conception of instruction as a triangle composed of interactions between and among teachers, students, and content. Professional development programs that include supporting curricular materials and formative assessment systems can target two parts of the triangle—teachers and content—thus increasing the chances for impacting instruction. 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 Writing Project sites—has been supporting teacher professional development since its origins as the Bay Area Writing Project in 1974. 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. 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
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, findings that have been replicated across multiple studies (Kim et al., 2011; Olson et al., 2012; Olson, Matuchniak, Chung, Stumpf, & Farkas, 2017). 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 implementation of CRWP. 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. The members of this team provided the intellectual leadership for the program, defining the student outcomes they sought to influence, the key program components, and the central support for local implementation (see Figure 1).

We begin with a discussion of student outcomes, then provide an overview of the technical assistance the National Writing Project’s CRWP leadership team designed to support local Writing Project sites to deliver a program that would lead to these outcomes. We then describe the key components of the program that sites signed on to implement with partner districts and the teacher practices they expected would come about through program supports.

(Figure 1 about here)

Student outcomes

At the outset of CRWP, the stated goal was to increase student writing proficiency in alignment with new college- and career-ready standards. With the new standards’ focus on informational and argument writing, CRWP was initially designed to develop instructional practice in both genres of writing. However, in Year 1, the CRWP leadership team began to place a greater emphasis on argument writing and, by the end of the first year, the team decided to focus on improving students’ abilities to write source-based arguments.
Technical assistance

From the outset, 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. As the leadership team gathered and made sense of information about local implementation, it adapted the supports it provided to sites.

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. They also facilitated the development of a community that led to cross-site sharing of knowledge and experience.

Members of the leadership team also assumed roles as “thinking partners”—experienced site leaders deeply versed in the program who had 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, used to determine how to support sites’ work.

During Year 1, the primary resource developed for the local Writing Projects was the Online Learning Experience, an online mini-course on argumentation and how to teach source-based argument writing. Each site sent to the class key “teacher consultants” (teacher leaders
connected to the local Writing Project sites) who were providing professional development in CRWP districts, 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 the research team.\(^1\) Through this process, the leadership team recognized 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 transfer ideas they were learning in the professional development into their classrooms. The curricular resources took the form of mini-units, developed by the current and former teachers on the CRWP leadership team, that systematically addressed key ideas in effective argument writing using evidence from nonfiction texts and became a focus of CRWP work in Year 2. 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 multiparagraph argument (see Figure 2).\(^2\) Additionally, by making explicit particular skills necessary for developing strong written arguments, the mini-units were constructed with the idea that they would be “educative” for participating teachers,

\(^1\) Inverness Research Associates partnered with SRI on the evaluation, conducting qualitative data collection and analysis and providing formative feedback to the CRWP leadership team.

\(^2\) See [https://sites.google.com/site/nwpcrwp/mini-units/](https://sites.google.com/site/nwpcrwp/mini-units/) for links to archived mini-units and related materials. The most commonly used mini-units were: Fast Food: Connecting Claims to Evidence, Reality TV, and Teen Brains. The CRWP leadership team has since refined and developed new mini-units. The most current mini-units at the time of our writing are available here: [https://sites.google.com/site/1516nwpcrwp/mini-units](https://sites.google.com/site/1516nwpcrwp/mini-units).
whose own content knowledge and pedagogical skills would grow through using them to teach.\textsuperscript{3} Moreover, unlike the Online Learning Experience in the first year, these resources directly reached both local Writing Project site teams \textit{and} teachers in CRWP districts.

\textit{(Figure 2 about here)}

The leadership team also developed the Using Sources Tool, a formative assessment that guided teachers through a series of prompts to analyze student writing (see Figure 3). 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. The idea was to focus teachers on specific aspects of argument writing (e.g., how student writing used evidence from sources to support the argument) and shift teachers’ emphasis away from the traditional focus on identifying and correcting grammatical errors in student writing. The formative assessment tool was also designed to be educative—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 \textit{Rewriting: How to Do Things with Texts}).

\textit{(Figure 3 about here)}

\textbf{Key program components}

With the program’s increased focus on argument writing and the additional resources and tools made available to sites, the key program components evolved from Year 1 to Year 2. Here we describe the key program components in both years, with a focus on Year 2. To assess implementation fidelity, the research team worked with the National Writing Project to establish implementation thresholds for each of the key program components; we note these thresholds in

\textsuperscript{3} 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).
the description here and refer back to them when reporting on implementation fidelity in the results section below.

**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 (at least 18 hours on argument and informational writing), and use strategies designed to support classroom implementation (at least 8 hours of 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 understanding of their district partners’ needs.

**Year 2.** As a result of the sharpened focus on source-based argument writing and the new resources designed to support classroom enactment, the National Writing Project expected local Writing Project sites’ implementation of CRWP to include the following program components, the first of which was unchanged from Year 1 to Year 2:

- **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 to emphasize support for classroom implementation. At least half of CRWP professional development events were to involve specific professional development formats: coteaching, demonstration lessons, coaching, coplanning and codesigning tasks, or analysis of student work.

- **Support for use of curricular resources.** Local Writing Project sites were expected to spend at least half their professional development events supporting teachers to learn about
and implement new curricular resources. The National Writing Project provided guidance to sites as to how to design this professional development (see Figure 4). Sites were also expected to ask teachers to implement four “replacement” mini-units in their classrooms over the course of the year.

- **Support for use of formative assessment tool.** In Year 2, CRWP also incorporated the regular use of the Using Sources Tool to focus teachers’ analysis of student work. Local Writing Project sites were asked to introduce their 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 requirement to submit ratings of student work created a sense of responsibility to ensure use of the Using Sources Tool and allowed for aggregated results to be shared at national convenings.

**Teacher instructional practices**

The National Writing Project designed the three key components of CRWP—the mini-units, the formative assessment tool, and professional development—to reinforce each other to support the development of teachers’ content knowledge and scaffold the transfer of new instructional ideas into teachers’ instructional practice. In particular, the mini-units were designed to catalyze the transfer of abstract ideas and new content knowledge into teacher instruction; the professional development provided critical scaffolds for rapid implementation; and the formative assessment positively reinforced teachers’ argument-writing instruction and helped teachers strategize about the skills to focus on in instruction. The National Writing Project anticipated that teachers’ instructional practices would shift as a result of this combination of supports. In particular, the goals were for teachers to (1) increase the frequency
with which they taught source-based argument writing and (2) increase their use of the approaches to argument-writing instruction that were embedded in the mini-units and reinforced by the Using Sources Tool, such as close reading and annotation of text, strategies for developing a claim and identifying supporting evidence, effective use of source materials, and use of writing processes (prewriting, drafting, revision).

**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 small size of the rural districts (from 1 and 24 ELA teachers in grades 7–10), randomization of entire districts allowed for the development of a community of practice among teachers that would not have been possible had randomization occurred at the school level. 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\(^4\) that were of a similar size, serving student populations with comparable demographics, 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 44 districts (22 pairs). In spring 2013, the research team randomized one district in each pair into treatment and control, allowing for professional development to begin in summer 2013. Treatment and control districts were of similar sizes and served observably similar student populations (see Table 1). The announcement of randomization decisions in spring 2013 did not differentially impact the observable characteristics of students.

(Table 1 about here)

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).\(^5\) They received a stipend of $2,500 per semester, which could be used for other educational purposes for teachers in grades 7–10 (e.g., textbooks, math professional development). After the 2 years of the evaluation, local

---

\(^4\) 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.

\(^5\) 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.
Writing Project sites provided a streamlined, 1-year version of CRWP to the control districts. No districts left the study after randomization.

**Teacher sample**

CRWP called for Local Writing Project sites to serve all grade 7–10 ELA teachers in CRWP districts. The study’s teacher population included all ELA teachers in grades 7–10 who taught a core ELA class (i.e., grade-level English class, excluding electives and support classes) at baseline (spring 2013) or for at least half a year during either study year. At baseline, the study’s teacher population included 329 teachers. By the end of Year 2, 43% (or 143) were no longer teaching grade 7–10 core ELA classes in the study districts. Treatment districts saw 48% attrition; control districts experienced 40% attrition. Of the teachers who left the study population, the majority left the district; approximately 10% changed assignment within the same school or district (e.g., new grade level, instructional coach). As new teachers took the place of departing teachers, these “in-mover” teachers joined the study’s teacher sample.

**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).

Given the expense of scoring student writing, we could not score all papers from the study population of students. Instead, we aimed to score an analytic sample of pre- and posttest writing from 60 students per district, which we estimated would provide a minimum detectable

---

6 The study excluded teachers in schools serving exclusively nontraditional students (e.g., alternative and continuation schools) or those without ELA departments (i.e., shared-time vocational schools).
effect size of .20 or below with 80% power. To provide an unbiased sample and provide an
estimate of student attrition, we randomly sampled students from the pretest student population
into an assigned sample without regard to whether they completed posttest writing. The analytic
sample therefore provides an unbiased sample of those students who remained in the district (and
completed the both pre- and posttests) between the beginning and end of the program.

To achieve the desired analytic sample size in the summer of 2015 (after the posttest), we
added to the assigned sample until we had 60 students per district in the analytic sample (i.e.,
with both pretest and posttest student writing). The students in the assigned sample without
posttest writing were considered attriters:

\[
\text{Attrition} = \frac{N_{\text{assigned\_sample}} - N_{\text{analytic\_sample}}}{N_{\text{assigned\_sample}}}
\]

This approach should leave an analytic sample of 1320 in each treatment and control (22
districts x 60 students). However, 5 of the 44 districts did not have 60 students with complete
pre- and posttest writing (primarily because of their small size), leaving an analytic sample of
1259 treatment students and 1227 control students. To achieve this analytic sample, we had to
sample 1791 students in the treatment group and 1860 students in the control group into the
assigned sample. The study’s student attrition is therefore 32% overall, 30% for treatment
students, and 34% for control students.

\[7\] Upon receiving completed pretest writing in fall 2013, we assigned a random number to all students for whom we collected these data, determining their order in the assigned sample. In the five districts with fewer than 60 students eligible for the analytic sample, we included all students with pre-test writing in the assigned sample and, by extension, all students with both pre- and posttest writing in the analytic sample. In the other 39 districts, we used the random number assigned at pre-test to sample students into the assigned sample until we had an analytic sample of 60 students per district, stratifying samples within districts into cells by district, baseline grade, and post-test prompt type to control for any potential prompt or grade effects. More information on the sampling strategy is available from the authors by request.
Inclusion of in-moving teachers and students

The analytic samples of teachers and students 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 spring 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 population of students eligible for sampling) were not collected until fall 2013. Given the high turnover expected in high-needs districts, the study included in-moving teachers for the sake of collecting data that accurately represented the instructional practices in each district throughout the study. In-movers could have been a threat to randomization if students or teachers changed districts to take advantage of the treatment being offered. 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. Although the particular circumstances of this project minimize the potential for bias by in-moving students and teachers (we randomized at the district level in rural communities, so parents’ choices to live and teachers’ decisions to work in one district over another are unlikely to be influenced by the treatment), we discuss this possibility in greater depth below.

Data and Methods

Over the 2-year study, researchers collected data to document (1) CRWP program implementation (professional development monitoring, 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 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 (format, content, and duration) and to record teacher participation. Sites submitted the forms three times per year (at the end of summer, fall, and spring semesters).

**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. 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 these 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 between .83 and .96; Gallagher et al., 2012), with modifications for better alignment with the instructional practices targeted by CRWP. Researchers administered the online instructional log once a day over 2 weeks in fall and spring of Year 2. Researchers administered the online teacher survey in spring
of Year 2. Administrations were timed to avoid 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 that day. The survey measured broader practices and constructs more appropriately measured over a year than in a single day. Teachers were asked 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 received an additional series of questions about 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 (AWC-SBA).

Over a decade, the National Writing Project developed the Analytic Writing Continuum (AWC), which has been shown to be a valid and reliable measure of student writing (Bang,
The original version of the AWC 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 Smarter Balanced). The National Writing Project worked with a panel of writing assessment experts to modify the AWC 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 AWC 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 AWC-SBA retains a basic structure rooted in the “6+1 Traits” of writing (Culham, 2003) but has a particular focus on the attributes related to source-based argument writing. The AWC-SBA measures four attributes:

1. Content (Including Quality of Reasoning and Use of Evidence): The content attribute describes how effectively the writing presents an argument supported by reasoning and developed through the use of evidence from sources.

2. Structure: The structure attribute describes how effectively the writing establishes an order

---

10 For more information about the National Writing Project Analytic Writing Continuum, see http://www.nwp.org/cs/public/print/resource/3776

11 For additional information on the development of the rubric, procedures to ensure unbiased scoring, the recruitment and training of scorers, and the scoring events themselves, see online Appendix A.
and arrangement to enhance the central argument.

- **Stance:** The stance attribute communicates a perspective through tone and style appropriate for the purpose and describes how effectively the writing establishes credibility.

- **Conventions:** The conventions attribute describes how effectively the writing demonstrates age-appropriate 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 deidentified the samples by removing names or other personal identifying information and assigning an anonymized identification number. The deidentified papers were sent to the National Writing Project for scoring. Scorers therefore did not know the district the papers came 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 provides descriptive statistics of student writing achievement on each of the AWC-SBA’s attributes by treatment condition. Note that, due to the sampling of equal numbers of students in each district (as available), student and district-level means are similar in size.

*(Table 2 about here)*

On the six-point rubric for the content attribute, 80% of papers received score of 2 to 4 on content; we provide excerpts of exemplars of these score points in Figure 5. Papers scoring a 1

---

12 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).
(9% of papers for the content attribute) were typically too brief to evaluate, consisted primarily of copied text, or lacked any discernible central idea (e.g., haphazard content). About 10% of the papers received a 5 or a 6 for content. These papers competently (score 5) or effectively (score 6) demonstrate reasoning by selecting and using evidence from sources to support the claim and including commentary on that evidence. Note that although papers received a different score for each attribute, the scores overall reflect a high degree of internal consistency (see Bang, 2013 for a discussion of internal consistency using the AWC).

(Figure 5 about here)

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. 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 model included a control for student
baseline achievement in fall of Year 1. 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. The models 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 district and block level (Kenward & Roger, 1997; Schaalje, McBride, & Fellingham, 2002).

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 results by 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 (days per teacher). To compare binary outcomes, we used melogit in Stata version 14.1. For ease of interpretation, 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. Note that tables providing data from teacher logs and surveys reproduce questions and Likert scale options from these instruments in their entirety.

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, we found that large majorities of teachers in nearly every district experienced a program defined by key program components in Year 1:

- In 20 of 22 districts, 80% or more of ELA teachers participated in 45 or more hours of CRWP professional development.
- In 20 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 21 of 22 districts, 80% or more ELA teachers participated in 8 or more hours of professional development designed to support classroom implementation.

Twenty districts reached all three of these thresholds, one district missed all three. and another district missed two.

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 support for use of curricular resources and a formative assessment tool.
Intensive professional development designed to support classroom implementation. As in Year 1, 20 of 22 CRWP districts met the implementation fidelity threshold for participation (80% of ELA teachers participating in 45 or more hours). Because of instability in the teacher workforce, just half the districts (11 of 22) reached the threshold for participation over both years (90 hours or more over 2 years).

In Year 2, the CRWP leadership team expected half of the CRWP professional development events to be delivered by using approaches that 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 via one of these formats). Demonstration lessons and coteaching were particularly common: nearly three-fourths (73%) of teachers reported observing Writing Project staff model the use of text-based argument tasks (i.e., CRWP mini-units) and/or benefiting from coaching or coteaching support from the Writing Project as they implemented CRWP tasks.

Curricular resources. In Year 2, the CRWP leadership team expected half of CRWP professional development events to support the use of CRWP curricular resources designed to teach argument writing from nonfiction sources. Fifteen of 22 districts met this threshold (for 80% of district ELA teachers, at least half of their CRWP professional development was on supporting the use of CRWP curricular resources). Per the guidance provided by the National Writing Project (see Figure 4), the professional development typically began with the local Writing Project introducing the mini-units, with the teachers experiencing the mini-units as if they were students and then analyzing and discussing their experiences. This opening was typically followed by collaborative planning focused on how to implement the units in teachers’
own classrooms: 89% of treatment teachers reported that planning with the Writing Project and/or colleagues supported their use of CRWP resources. Planning often involved thinking through the role that the Writing Project professional development provider would play in supporting classroom implementation—for example, in some cases the professional development provider introduced the mini-unit as a demonstration lesson; in others, the professional development provider and the teacher planned to coteach a lesson or two. Ninety-three percent of ELA teachers in CRWP districts reported having taught at least one CRWP mini-unit.

**Formative assessment.** In Year 2, the CRWP leadership team asked local Writing Project sites to introduce ELA teachers to the CRWP 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. 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. Overall, 93% of ELA teachers in CRWP districts reported analyzing student writing with the Using Sources Tool. In nearly all cases, teachers analyzed student work collectively during a professional development event. This use of the formative assessment tool provided teachers with rapid feedback on student learning that was tightly aligned with the desired student outcome. This feedback came as they were implementing new ideas and suggested next steps in instruction, thus providing critical data to reinforce motivation for uptake and information to guide implementation.

Looking across the five indicators of the key program components (putting aside the indicator for cumulative participation over 2 years, see Figure 1), two districts fell short of meeting two indicators, eight districts missed a single indicator, and 12 districts met all
The survey revealed large, statistically significant differences between teachers’ experience with professional development in CRWP and control districts. In 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%) (see Table 3). 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 analyzing student work.

(Table 3 about here)

**Impacts on teacher practice in CRWP districts**

As a result of CRWP, teachers’ practice changed in focus, with teachers using writing time to teach the elements of argument writing emphasized by CRWP. Table 4 provides outcomes from the 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) 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 practicing argument.

(Table 4 about here)
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. In addition, CRWP 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. Although 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. Both site visit observations and survey findings suggest that much of the teaching being replaced was on conventions and literary analysis. Control district teachers reported a greater emphasis on teaching conventions (4.0 vs. 3.6 on a 5-point scale), and they were nearly four times as likely as
CRWP teachers to report that the most intensive writing assignment given to students was in literary analysis (19% vs. 5%). 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, but it would change the nature of that instruction.

*(Table 5 about here)*

**Impacts on quality of student writing in CRWP districts**

CRWP had 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 marginally significant (*p* < .1). Estimates of impacts on content, structure, and stance are robust to alternate model specifications (Table 7).

*(Tables 6 & 7 about here)*

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 that 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 writing has been identified as critical to college and career readiness and thus is central to new state standards for English language arts. Moreover, the National Writing Project demonstrated these outcomes at a large scale (22 districts in 10 states) and in challenging settings (rural high-poverty secondary schools).

Given the inclusion of in-moving teachers and students, there remains a possibility that these results could be driven not by changes on student achievement or teacher practice, but rather by the treatment serving as an effective recruitment tool for attracting higher-achieving students and teachers to the districts. We argue that this concern, especially given districts as the unit of assignment, is unlikely to have actually driven these results. We announced randomization decisions in the spring prior to Year 1 and selected the study student population from students who were present in their district in fall of Year 1, prior to any changes in teacher practice or district achievement. Baseline equivalence shows no differences in student writing achievement between the treatment and control districts, and observable student characteristics remained equivalent between the two sets of districts in the years before and after randomization decisions were announced. Finally, the combination of overall and differential attrition from
these baseline samples is low enough that it is unlikely to introduce bias (U.S. Department of Education, 2014).13

The overall teacher attrition rate was 43% (high levels of teacher turnover are common for rural communities, see Monk, 2007) and the differential attrition rate was 8%, exceeding the liberal standard for differential attrition in randomized-controlled trials. It is possible that lower performing teachers may have left treatment districts because of CRWP and higher performing teachers may have been attracted to the program, this assumes that a meaningful number of such teachers based decisions about their employment on their interest (or lack thereof) in a particular professional development program. This cannot be dismissed but it seems unlikely that, at this magnitude, differential attrition or in-moving teachers drive the study results. A future study will examine this in greater detail.

Perhaps more notable is that CRWP districts saw positive effects despite the fact that roughly one-third of teachers received only 1 year of the 2-year program. In some prior randomized controlled trials of professional development programs, the issue of teacher turnover has been posited as a likely cause of no-effects findings (Desimone & Garet, 2015; Garet et al., 2011; Garet et al., 2008).14 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 ways to influence instruction that could impact student

13 Since randomization occurred at the district level in rural communities, to believe that differential recruitment of students drove these results, one would have to posit that parents of observationally similar students with greater growth potential moved their families to a new community to take advantage of a single-subject teacher professional development program before this program had demonstrated its effectiveness in these district. Additionally, we would have to believe that enough of these parents moved to bias the results and that this differential movement occurred only because of the announcement of the randomization decisions but stopped quickly thereafter. The implausibility of this combination suggests the estimated CRWP impacts reflect increased achievement of students and these results are not biased by in-moving students.

14 Desimone and Garet (2015) describe teacher turnover as a challenge facing urban schools. CRWP data also serve as a reminder that it is also a challenge in high-poverty rural districts.
learning. 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 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 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 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.

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—can lead to meaningful student learning. The findings therefore have important implications for practitioners, policymakers, and program developers.
References


Context for Teaching and Learning

- Policy context (college- and career-ready standards, assessment 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 writing
- 45 hours per year
- 80% of grade 7–10 ELA teachers participate
- Use of formats that support classroom implementation (e.g., coaching, demonstration lessons)

**Support use of curricular resources (Y2)**
- Introduce argument writing mini-units
- Support classroom implementation of mini-units

**Support use of formative assessment tool (Y2)**
- Introduce formative assessment tool
- Collaboratively use tool to examine writing
- Use analysis to inform next steps for professional development and teacher practice

Teacher Instructional Practices

**Teaching argument writing with nonfiction sources**
- Close reading and annotation of nonfiction text
- Development of claim and evidence
- Effective use of source material
- Writing processes (prewriting, drafting, and revision)

Student Outcomes

**Proficiency with source-based argument writing**
- Writing presents a claim supported by a clear line of reasoning and developed through the use of evidence from sources
- Establishes credibility of source material and includes commentary connecting evidence from sources to the line of reasoning
Figure 2. Key Features of a CRWP Mini-Unit

- **A text set**—At the center of each mini-unit is a non-fiction text set that is focused on a topic of interest to teens, designed to connect students to topics that will invite them to write. The texts vary in purpose, length, and format, with texts chosen specifically to provide multiple perspectives on a particular issue.

- **Close reading and exploratory writing**—Embedded strategies for reading and comprehending the texts are detailed so that students slow down enough to think about the facts, issues, and perspectives involved. Guidance to support students in identifying evidence that could be used to support a claim is included among the resources.

- **Argumentation focus**—Each mini-unit emphasizes at least one particular element of argumentation (i.e., a skill or writing move that helps students make effective arguments). The intent behind the series of mini-units is to work more intensely on a particular aspect of argument writing, master it, and then take up another mini-unit that will focus on a different, but equally important move that argument writers make.

- **Writing processes**—The mini-units call for students to draft their own texts and revise them after feedback from peers and/or the teacher. Each mini-unit suggests specific ways to do this in the course of a short unit.

Source: Adapted from CRWP’s “Moving Mini-Units Into the Classroom” resource for site leaders

Figure 3. The CRWP Using Sources Tool

The CRWP Using Sources Tool walks teachers through a series of questions designed to support their analysis of student work. These questions include:

- How would you describe the writing’s use of source material? (*Response options were on a 5-point scale ranging from “skillfully integrates” to “does not use source material.”*)
- Does the writing distinguish between the student’s own ideas and the source material? (*Response options were on a 4-point scale ranging from “effectively” to “not present.”*)
- Does the writing comment on the source material? (*Response options were on a 4-point scale ranging from “effectively” to “not present.”*)
- Does the writing characterize the credibility of the source material or the author? (*Response options were on a 4-point scale ranging from “effectively” to “not present.”*)
- Does the writing use source material for any of the following purposes? (*Mark all that apply.*)
  - Illustrating—use specific example from the text to support the claim
  - Authorizing—refer to an “expert” in the claim
  - Extending—put your own spin on terms and ideas taken from other texts
  - Countering—“push back” against the text in some way (e.g., disagree with it or interpret it differently)
- What do you see as next steps for this student? (*Open ended*)
Figure 4. Design for Professional Development in Support of Mini-Units

The CRWP leadership team encouraged the following professional development sequence designed to support implementation of the mini-units:

- Teachers experience a mini-unit as a writer
- Teachers analyze the process as educators (What is being taught? Learned?)
- Teachers add any scaffolding or alternate text necessary for student success
- Teacher consultants model in teachers’ classrooms
- Teachers present with teacher consultant coaching at least once during the mini-lesson series
- Teachers implement the rest of the mini-unit independently with regular check-ins by TCs via phone or email
- Teacher consultants and teachers meet to reflect on implementation (CRWP provides questions for discussion): share impressions, examine student work samples using formative assessment tool and teacher/student reflections; learn what students understand about argument writing
- Decide what to teach next
Figure 5. Excerpts from Student Source-based Argument Writing

<table>
<thead>
<tr>
<th>Exemplars of writing scoring a 2, 3, and 4 on the content attribute illustrate the differences between these three most common score points. Scorers assessed for several facets of content to assign scores, including reasoning, development and selection of evidence, commentary, alternate perspectives, and signaling the use of source material. Below we excerpt illustrations of commentary on source material at each of these points. Note that these prompts were written on demand and without time explicitly provided for revision. Excerpts are provided with original grammar and spelling; both the levels of sophistication and correctness were reflected in the conventions, not content, score. Papers provided are exemplars of score points for the content attribute and may have scored differently on the other three attributes. The papers are reproduced in their entirety in the online Appendix B. A paper scoring 2 on the content attribute will likely present some information from source material, but it may be directly copied and will rarely be commented upon (e.g., elaborated upon or interpreted). In the excerpt below, we see this score-point 2 exemplar include relevant detail from the source material on Title IX, though the writing does not comment on the ideas presented:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Famous people have put an end to sexist sports. Paul Bucha, a retired Army colonel who worked with Ike to plan the Normandy invasion, sued to help his daughter become a champion swimmer. Former big league pitcher, and future manager, Dallas Green lead to have no more only little League’s boys-only.</strong></td>
</tr>
<tr>
<td>A paper scoring 3 on the content attribute typically includes minimal commentary on source material. Below, we see the exemplar score-point 3 paper elaborate on the source material by asking—and answering—a rhetorical question. This minimal commentary is repeated at the end of most body paragraphs:</td>
</tr>
<tr>
<td><strong>Title IX has allowed all women equal opportunity to follow their dreams... Although this is true some say that it has taken away from mens sports and mens opportunity. In some cases this is true, but ask yourself this. Is sacrificing a portion of mens sports/opportunity worth all women opportunity. I believe so.</strong></td>
</tr>
<tr>
<td>A paper scoring 4 on the content attribute will satisfactorily present limited commentary on source material. For example, in the excerpt below, the writing uses data from the source material to elaborate on the argument that prisoners’ organs are safe to donate:</td>
</tr>
<tr>
<td><strong>Some that are against prisoners donating say that prisoners are not yet healthy enough or are bad organs. Yet in reality, “the government solicits donors from communities with equally high or higher disease rates than prisoners.” The worry about disease in first place is totally unnecessary. Today medial advancements have made It to where screening a prisoner Is accurate 99% of the time. So the ones that do have disease, we will know.</strong></td>
</tr>
</tbody>
</table>
### Table 1. District Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students</td>
<td>1,881</td>
<td>1,864</td>
</tr>
<tr>
<td>Free or reduced-priced meals</td>
<td>65%</td>
<td>68%</td>
</tr>
<tr>
<td>Students of color</td>
<td>36%</td>
<td>37%</td>
</tr>
<tr>
<td>English learners</td>
<td>3%</td>
<td>3%</td>
</tr>
</tbody>
</table>


### Table 2. Student Outcome Descriptives

<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><strong>Content</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student mean</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Student s.d.</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>District mean</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>District s.d.</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student mean</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Student s.d.</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>District mean</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>District s.d.</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>Stance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student mean</td>
<td>2.7</td>
<td>2.6</td>
</tr>
<tr>
<td>Student s.d.</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>District mean</td>
<td>2.7</td>
<td>2.6</td>
</tr>
<tr>
<td>District s.d.</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td><strong>Conventions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student mean</td>
<td>2.7</td>
<td>2.8</td>
</tr>
<tr>
<td>Student s.d.</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>District mean</td>
<td>2.7</td>
<td>2.8</td>
</tr>
<tr>
<td>District s.d.</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Student n</td>
<td>1,259</td>
<td>1,227</td>
</tr>
<tr>
<td>District n</td>
<td>22</td>
<td>22</td>
</tr>
</tbody>
</table>
Table 3. Differences in Writing Professional Development Received by Treatment Status, Year 2

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th>Control</th>
<th>Block n</th>
<th>District n</th>
<th>Teacher n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Since the end of last school year, how many hours of PD on writing instruction have you received?</td>
<td>63.41</td>
<td>6.41 ***</td>
<td>22</td>
<td>44</td>
<td>301</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>22</td>
<td>42</td>
<td>214</td>
</tr>
<tr>
<td>Received PD focused on writing from source material (1=Yes)</td>
<td>91%</td>
<td>50% ***</td>
<td>22</td>
<td>42</td>
<td>214</td>
</tr>
</tbody>
</table>

**To what extent did the PD involve...**

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

To what extent do you dis(agree): I received adequate professional development to teach writing**

<table>
<thead>
<tr>
<th>Scale: 1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree</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>4.27</td>
<td>2.97 ***</td>
<td>22</td>
<td>44</td>
<td>305</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 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.
Table 4. Differences in Daily Student Writing by Treatment Status, Year 2

<table>
<thead>
<tr>
<th></th>
<th>Treatment</th>
<th></th>
<th>Control</th>
<th></th>
<th>Block n</th>
<th>District n</th>
<th>Teacher n</th>
<th>Log n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you ask students to write? (1 = Yes)</td>
<td>94%</td>
<td></td>
<td>89%</td>
<td></td>
<td>22</td>
<td>44</td>
<td>341</td>
<td>2,938</td>
</tr>
<tr>
<td>How long did they write (in minutes)?</td>
<td>32.66</td>
<td></td>
<td>31.11</td>
<td></td>
<td>22</td>
<td>44</td>
<td>334</td>
<td>2,382</td>
</tr>
<tr>
<td>How much did they write?&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.80</td>
<td></td>
<td>2.50</td>
<td></td>
<td>22</td>
<td>44</td>
<td>332</td>
<td>2,305</td>
</tr>
</tbody>
</table>

<sup>When asked to write, did student work include:</sup>

<table>
<thead>
<tr>
<th>Activity</th>
<th>Treatment</th>
<th></th>
<th>Control</th>
<th></th>
<th>Block n</th>
<th>District n</th>
<th>Teacher n</th>
<th>Log n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice writing an argument</td>
<td>26%</td>
<td></td>
<td>7%</td>
<td></td>
<td>22</td>
<td>44</td>
<td>334</td>
<td>2,463</td>
</tr>
<tr>
<td>Generate an argument in support of a claim</td>
<td>26%</td>
<td></td>
<td>10%</td>
<td></td>
<td>22</td>
<td>44</td>
<td>334</td>
<td>2,463</td>
</tr>
<tr>
<td>Identify a claim and evidence in text</td>
<td>27%</td>
<td></td>
<td>26%</td>
<td></td>
<td>22</td>
<td>44</td>
<td>334</td>
<td>2,463</td>
</tr>
<tr>
<td>Develop a claim</td>
<td>36%</td>
<td></td>
<td>17%</td>
<td></td>
<td>22</td>
<td>44</td>
<td>334</td>
<td>2,463</td>
</tr>
<tr>
<td>Evaluate the credibility of evidence</td>
<td>19%</td>
<td></td>
<td>10%</td>
<td></td>
<td>22</td>
<td>44</td>
<td>334</td>
<td>2,463</td>
</tr>
<tr>
<td>Support a claim with evidence from a text</td>
<td>44%</td>
<td></td>
<td>41%</td>
<td></td>
<td>22</td>
<td>44</td>
<td>334</td>
<td>2,463</td>
</tr>
<tr>
<td>Support a claim with evidence from personal experience</td>
<td>13%</td>
<td></td>
<td>6%</td>
<td></td>
<td>22</td>
<td>44</td>
<td>334</td>
<td>2,463</td>
</tr>
<tr>
<td>Elaborate upon evidence used to support a claim</td>
<td>24%</td>
<td></td>
<td>15%</td>
<td></td>
<td>22</td>
<td>44</td>
<td>334</td>
<td>2,463</td>
</tr>
</tbody>
</table>

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.

<sup>a</sup> 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.

*<sup>p</sup> < .05. **<sup>p</sup> < .01. ***<sup>p</sup> < .001.
## 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?**

<table>
<thead>
<tr>
<th>Argument</th>
<th>82%</th>
<th>39%</th>
<th>***</th>
<th>22</th>
<th>44</th>
<th>310</th>
</tr>
</thead>
<tbody>
<tr>
<td>Literary Analysis</td>
<td>5%</td>
<td>19%</td>
<td>***</td>
<td>22</td>
<td>44</td>
<td>310</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.
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>
</tbody>
</table>

N's

<table>
<thead>
<tr>
<th></th>
<th>Block</th>
<th>District</th>
<th>Student</th>
</tr>
</thead>
<tbody>
<tr>
<td>N's</td>
<td>22</td>
<td>44</td>
<td>2,486</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>22</th>
<th>44</th>
<th>2,486</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>df, Impact estimate</th>
<th>21</th>
<th>21</th>
<th>21</th>
</tr>
</thead>
</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>Content</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structure</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stance</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conventions</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

Random block effects X

Fixed block effects

Student-level prior achievement control X

District-level prior achievement control X

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.
Appendix A: Scoring Student Writing

Student responses to the on-demand writing prompts were scored using the National Writing Project’s Analytic Writing Continuum for Source-Based Argument (AWC-SBA) rubric. SRI independently collected all student writing, removed any identifiable information from the papers to blind the writing to treatment and control condition and time point, and monitored NWP’s scoring of the blinded papers. Scoring occurred during a single scoring conference held in summer 2015. Scorers applied the AWC-SBA to assess four different attributes of student writing—content, structure, stance, and conventions—on a six-point scale. To ensure that teachers did not align their instruction to the outcome measure, neither CRWP nor control districts had access to the AWC-SBA rubric. The local Writing Project sites that facilitated the CRWP professional development also did not have access to the AWC-SBA rubric or any supporting materials. Additional detail on the preparation of papers, recruiting and training of scorers, and scoring process is provided below.

Development of rubric. For prior studies, the National Writing Project has used the Analytic Writing Continuum (AWC) as an outcome measure. The AWC was originally based on the 6+1 Trait Writing Model (Culham, 2003) and has been used by the National Writing Project for over a decade to measure writing quality across genres and grade levels and has demonstrated high reliability across multiple past National Writing Project scoring events (Bang, 2013). Given CRWP’s explicit focus on source-based argument, the National Writing Project convened a panel of national experts to adapt the AWC to measure genre-specific writing (e.g., development of reasoning, signaling that ideas are taken from a source). The resulting rubric, the AWC-SBA, was winnowed to four attributes, described in the main text. As a small number of students (168,
or 7% of the analytic sample) wrote source-based informational texts at baseline, the National Writing Project also created and piloted a rubric aligned to source-based informational writing, the Analytic Writing Continuum for Source-Based Informational Writing (AWC-SBI). The National Writing Project refined these rubrics in a pilot scoring in spring 2015.

To support scorers in using the rubrics reliably, the National Writing Project convened an expert panel of 14 writing assessment experts, 10 of whom were independent consultants, to select anchor papers for each score point for each rubric and grade span (7–8 and 9–10). Anchors were selected from on-demand student writing from study districts that would not be scored for the baseline or outcome data. This panel also wrote commentary to describe the rationale for assigning each anchor paper the scores it received. The “exemplar” papers excerpted in the main text (and provided in their entirety in Appendix B) are anchor papers for the high school AWC-SBA scoring.

**Assurance of nonbiased scoring.** SRI collected student writing from each study district with the help of a local research site coordinator (LSRC), who lived near the study districts and facilitated all local data collection (e.g., collection of teacher rosters, distribution and collection of writing prompts). Local Writing Project sites assisted in the recruitment of LSRCs, who were managed and paid by SRI. LSRCs were not involved in any way the CRWP professional development.

To prepare the sampled students’ responses to the writing prompts, SRI redacted all identifying information and tagged prompts with a unique ID. To de-identify the student writing, SRI removed any personally identifiable information to protect student privacy and any information about experimental condition (e.g., district, school, teacher) or time point (e.g. Appendices
baseline, Year 2 outcome) that might bias a scorer’s rating. SRI tagged each writing sample with an individual ID system using a format created by the National Writing Project to facilitate accurate movement of papers through the scoring process (e.g., the ID contained information used to ensure that student papers that were randomly sampled by SRI for double-scoring got double-scored). The use of two ID systems ensured that neither scorers nor National Writing Project staff had access to information about which writing samples were drawn from treatment versus control districts. SRI used a cross-walk file to link each paper’s scoring ID to the more detailed ID that SRI used for tracking and analysis.

**Recruiting and training scorers.** To score the student writing, the National Writing Project recruited teachers affiliated with local Writing Project sites that were not involved in CRWP. All scorers had a minimum of 3-years experience teaching writing either in the grade levels or an adjacent grade to the grade levels they were scoring. This recruitment strategy ensured that teachers were knowledgeable about grade-level appropriate expectations for writing development. Where possible, the National Writing Project recruited scorers with previous scoring experience. Approximately one-third of scorers had previously scored papers using the AWC. However, the emphasis on narrative writing in the curriculum and newly developed genre-specific rubrics meant scorers generally had limited experience scoring source-based argument (or informational) writing. 105 scorers participated in the ABC-SBA scoring; 10 scorers remained for the AWC-SBI scoring that immediately followed the AWC-SBA scoring.

Writing assessment experts, who participated in the selection and annotation of anchor papers, led the scorer training with the assistance of table leaders, who oversaw scoring of 4 to 5 scorers. (None of the scoring or table leaders were National Writing Project staff members.) Table leaders were recruited from participants from the pilot scoring. National Writing Project
leaders provided an initial training to the table leaders before scorers arrived. This initial training consisted of leadership reviewing the scoring rubrics and walking through a set of grade-level specific anchor sets and calibration papers (selected from the sample of papers collected by the CRWP evaluation but not sampled for scoring). In addition to its focus on the rubric, the training covered the logistics of scoring.

The National Writing Project leaders then trained the larger group of scorers with support from the table leaders. For training, scorers separated into a 7th–8th grade room and a 9th–10th grade room to calibrate scoring to grade-level appropriate standards. The training focused on the interpretation and application of the scoring rubrics. Scorers had extended time to practice scoring sample papers that had previously been scored by the panel of writing assessment experts and selected by the scoring leaders. The results of the calibration session were monitored by the scoring leaders to ensure that scorers were within one-point agreement of the scores assigned by the panel of writing assessment experts.

**Conducting the scoring sessions.** Student responses to writing prompts collected in both fall 2013 and spring 2015 prompt administrations were scored together. Tables of scorers would read papers written on a single topic (e.g., Title IX) during each 90-minute scoring session. Each scoring session would begin with scorers reading the source material provided to students during the on-demand task. Research staff distributed writing samples from the different districts randomly across the scoring tables. Following every major break (e.g., overnight, after meals), all scorers recalibrated using practice sample papers before operational scoring resumed. Table leaders “read behind” scorers at their tables during the actual scoring session, monitoring individual scorers’ application of the rubrics and consulting with or retraining scorers when warranted.
For the 15% of double-scored papers, if two raters assigned scores that were one point apart, the scores were averaged. If the raters assigned scores that were two points apart, a table leader adjudicated the scoring by independently scoring the paper; the leader’s score was assigned to the paper.

The de-identification described earlier ensured scorers were “blind” to the school’s experimental condition and the random distribution of papers across scorers and scoring sessions ensured that any error in scoring (i.e., rater effects or drift during the scoring) would be randomly distributed across districts. Any error in scoring would therefore inflate the standard error of the impact estimate, but would not bias the point estimate measuring the size of the treatment effect.
Appendix B: Sample Student Writing

In this appendix, we provide the complete and unedited examples of student writing from which samples were excerpted in the main text. As explained in Appendix A, the National Writing Project used these writing samples as anchor papers to illustrate exemplars of scoring points 2–4 in the scoring of all baseline and outcome papers.

The first paper serves as an anchor for the following score points: content 2, structure 2, stance 2, conventions 3.

*I think Title IX should stay unchanged because woman’s sports are just as important as mens sports. People should not be sexist about sports. It is unfair to change woman’s sports because they are also just as physical as men’s sports. For example Pat Summit, in 38 years, had won 8 national championships and more than 1,000 wins, having more than any other coach out there. “Every single woman who ever played for Pat has either graduated or is on her way to a degree” (Obama 3). One person has filed nearly 1,000 Title IX discrimination complaints, had been involved in more than “62.5% of the actions brought before the Office for Civil Rights in the state of Washington” (Alexander 7).

The 1970’s feature a few unlikely Title IX champions. Famous people have put an end to sexist sports. Paul Bucha, a retired Army colonel who worked with Ike to plan to Normandy invasion, sued to help his daughter become a champion swimmer. Former big league pitcher, and future manager, Dallas Green lead to have no more only little League’s boys-only.

In conclusion people shouldn’t change Title IX.

The second paper serves as an anchor for the following score points: content 3, structure 3, stance 4, conventions 3.
Title IX has been very helpful to all women. It has allowed all women equal opportunity to follow their dreams. Title IX has influenced all women to become leaders and fill important roles in society. Title IX has also influenced decision making among women, it is a fact women that play sports or go to college are less likely to do drugs. So Title IX has clearly been influential in a good way to our country and women all across America.

Title IX has allowed all women equal opportunity to follow their dreams. It has done so because it says that all women have the right to go to college and or pursue their dreams of playing college or professional sports. Allthough this is true some say that it has taken away from mens sports and mens opportunity. In some cases this is true, but as yourself this. Is sacraficing a portion of mens sports/opportunity worth all women opportunity. I belive so.

Another topic that must be brought up in the addressing of Title IX is that its not just for sports. Title IX supports all women in not just their right to play sports, but go to college as well. This statement brings me to the conclusion that this gives women the opportunity to become leaders and fill important roles to society. Barack Obama, in his passage supporting Title IX says that “women that grew up with it are now pioneering scientific breakthroughs, running successful businesses, governing states, and yes coach varsity teams. Now I ask this again, is a portion of mens opportunity/sports worth everything that helps us as a nation, I belive so.

Another thing that Title IX has helped women with is that, women who go to college or play sports are less likely to do drugs. Out of the many things Title IX does this is one of the most important. It has helped women to not do drugs, not make bad decisions, and above all it helps our nation thrive for success.

In conclusion, I belive Title IX has been extremely important and crucial to women and the nation as a whole. Whether you agree or dissagree with me, you mus admit that it has helped
and its good for everyone.

The third paper serves as an anchor for the following score points: content 4, structure 4, stance 4, conventions 3.

Dear Prision Board,

Donating apart of you to help another human Being is a big deal. Taking a organ out of ones body to put in another is crazy right? 14,275 are the amant of donors there are. A lot of people have Donated to people in need, and it has helped many people survive. There are people, as you know, that are not allowed to Donate, prisinors. I beilive that prisonors should be allowed to donate because the surgery is safe, they wouldnt be forced, and the could contribute alot.

To give a part of yourself, first you need to remove It from yourself. That’s where surgery comes In. The surgery required is a big procedure but as reading I says, “most donors remain healthy after the surgery. Some that are against prisinors donating say that prisonors are not healthy enough or are bad organs. Yet in reality, “the government solicits donors from communities with equally high or higher disease rates than prisonors.” The wory about disease in first place is totally unnecessary. Today medical advancements have made It to where screening a prisioner Is accurate 99% of the time. So the ones that do have disease, we will know.

Alot of people think that prisinors will be coerced into donating. They think that prisonors will expect a reward for donating. The solution is to make It completely voluntary. So that the prisoner has to want to do it, and Is informed that he/she will recive no benifits. This simple solution Is why prisonors should be allowed to Donate. Also they are already allowed to Donate to their family so whats the big deal If they donate to someone else instead.
If allowed prisoners could make a huge impact on the situation. There are currently 121,272 people waiting for an organ. There are only 14,275 Donors. Way too big of a difference. If prisoners were allowed just 5% of them could supply a kidney to everyone waiting. Just 5%. The willingness of the prisoners to donate is also present. “80% of people on death row said they would donate if they could.” Also thousands signed up for a previous Donation. The prisoners want to help. Why shouldn’t we let them.

So in conclusion, letting prisoners donate is a great idea. The prisoners would be fine if they did donate, they wouldn’t be forced. And the could and want to make an impact. If someone wants to help who are we to say no.