Abstract Title Page

Title: Impact Evaluation of the National Writing Project's College-Ready Writing Project in High Poverty Rural Districts

Authors and Affiliations: H. Alix Gallagher, Nicole Arshan, Katrina Woodworth
SRI International

Abstract Body

Background / Context:

Writing is an essential skill for participating in modern American society. Although it is crucial to careers and civic engagement, student writing falls far short of national expectations (College Board, 2004; NCES, 2012; Persky, Daane, & Jin, 2003). The Common Core State Standards (CCSS) seek to increase the rigor of writing instruction students receive and also to increase the proportion of instruction focused on students' abilities to write argument and informational text because of their importance for success in college and careers (Calkins, Ehrenworth, & Lehman, 2012; Rothman, 2011; Cutler & Graham, 2008).

With these new and loftier goals for student writing outcomes, practitioners and program developers are seeking guidance on how to help teachers work with students to obtain them. Teacher professional development is a logical approach. Recent findings from large-scale random assignment studies focused on impacting student outcomes via teacher development are mixed at best (Yoon, et al 2007; Garet et al, 2008; Garet et al, 2011), providing little experimental evidence upon which to build effective programs to improve teaching quality. Well developed theory about instructional capacity (e.g., Cohen and Ball, 1999) suggests that student learning is produced by the interactions between teachers, students, and educational material. As a result, programs that combine professional development and aligned curriculum may be more likely to change teacher practice enough to improve student outcomes.

In this context, the National Writing Project (NWP) won an Investing in Innovation (i3) grant to provide professional development to secondary teachers in rural districts with the goal of improving teachers' ability to teach to Common Core writing standards and ultimately students' writing proficiency. This paper presents a district-randomized controlled trial that found positive impacts of the NWP's College-Ready Writers Program (CRWP) using a validated measure of students' text-based argument writing as the outcome measure.

Purpose / Objective / Research Question / Focus of Study:

This study had three main research questions:

1. Was CRWP implemented with fidelity?
2. What impacts did CRWP on teachers' instructional practice?
3. What impacts did CRWP have on student argument writing?

Setting:

The study was conducted in 44 rural, high-needs school districts in 10 states: Alabama, Arizona, Arkansas, Louisiana, Mississippi, Missouri, New York, Oklahoma, South Carolina, and Tennessee.

Population / Participants / Subjects:

The study sample consists of all ELA teachers in the 44 study districts who taught a core ELA class for students in grades 7-10. See Table 1 for district enrollment sizes and descriptive statistics by treatment status. The student study population included the students who began in the 7th–9th grades in the 2013–14 school year and were still enrolled in the 2014–15 school year. Researchers randomly sampled 20 students per grade per district and scored baseline and outcome data for these students.[1]

[1] Five districts had fewer than 60 students available to score; in these districts we sampled the population.

Intervention / Program / Practice:

The CRWP intervention is designed to improve student writing of argument text in grades 7-10 by improving teachers' instructional practices in writing. CRWP builds on the NWP's extensive experience providing intensive in-service to teachers in writing instruction. As with all NWP work, local university-based site affiliates provide professional development with support from their national office and peers. Three key components define CRWP professional development: the duration and breadth of teacher participation, professional development content and resources, and strategies for delivering professional development.

- *Duration and breadth of participation*. The CRWP model calls for at least 80% of 7th-10th grade ELA teachers in each district to participate in at least 90 hours of CRWP professional development over 2 years (or 45 hours per year); the idea is that, to significantly influence teachers' instructional practice across a district, a substantial proportion of the faculty must participate in an in-depth professional development experience.
- *Content and resources*. The content of the professional development is shaped by both the expectations of college- and career-ready standards and the research on high-quality teaching of writing. The professional development content focused on argument writing from non-fiction sources. Teachers were provided and supported to use (1) a wide range of curricular resources designed to support argument writing and (2) a formative assessment tool designed to evaluate students' use of source material in their writing.
- *Strategies*. CRWP focused on classroom enactment (via demonstration lessons, coaching, designing learning tasks, planning, and examining student work) to support teachers in regularly using the new curricula resources and formative assessment tools.

Research Design:

The evaluation used a cluster-randomized controlled trial design that randomly assigned 44 rural, high-need school districts to the treatment or control condition. Participating districts were recruited in pairs and then randomized within pair to balance contextual and state policy issues that impact the local region. The control condition was "business-as-usual." The study was designed to examine both the implementation and impact of the intervention and included data collection to understand the counterfactual/control condition.

Data Collection and Analysis:

The source for student outcome data is a study-administered student writing assessment, while teacher outcomes were measured via both a teacher survey and an instructional log. In addition, the study team examined fidelity of program implementation by collecting data on the intervention from the participating Writing Project sites. Finally, the study relied on the teacher survey as well as district site visits to understand the treatment-control contrast.

Study-administered on-demand assessments of source-based argument writing were given to all students in grades 7 through 9 as a baseline (fall 2013) and all students in grades 8 through 10 as an outcome (spring 2015) as a measure of student writing ability. The prompts provided students with 4-6 short texts to read and students were asked to respond by writing an argument using evidence from the texts. The prompts were scored according to the Analytic Writing Continuum—Source-based Argument (AWC-SBA), a rubric developed by the NWP originally based on the widely known "6 traits" of writing and adapted to measure source-based

argument writing, with a particular focus on quality of reasoning and strength of evidence.[2] The writing was scored on four attributes: content (e.g., quality of reasoning and strength of evidence); structure (e.g., organization); stance (e.g., tone, perspective, establishment of credibility), and conventions. See Table 2 for descriptive statistics on each attribute by treatment status.

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:

$$Y_{ijk}=\beta_0 + \beta_1(\text{Treatment}_j) + \mathbf{(Covariates_i)\beta_2} + \varepsilon_{ijk} + \eta_{jk} + \mu_k$$

Random effects ε_{ijk}, η_{jk}, and μ_k allow for error at the student, district, and block level, respectively. Vector $\mathbf{Covariates_i}$ provides controls for student baseline achievement and outcome prompt taken. Student baseline scores are standardized within cohort to account for both prior achievement and cohort at baseline. $\mathbf{Covariates_i}$ also includes a vector of dummies for outcome prompt administered by grade, centered within the analytic sample. β_1 provides an estimate of the effect of district assignment to treatment on student writing ability within that district (the Intent-to-Treat effect).

Teacher outcome measures included a daily instructional log and survey. The instructional log was administered once a day over three weeks, with a baseline week in spring 2013 and outcome weeks in fall and spring of the 2014–15 school year. The annual teacher survey was administered at baseline in spring 2013 and again as an outcome measure in spring 2015. To estimate the impact of CRWP on teacher outcomes, we compare survey and log indicators of teaching practice using similar methodology to that used to analyze student-level outcomes. Survey data also require a three-level HLM (blocking pair, district, and teacher), whereas the logs require a fourth level in the models (the logs themselves), given the multiple measurements per teacher.

To measure the fidelity of program implementation, the study team collected data from local Writing Project sites on the three key components of the professional development: duration and breadth of participation, content and use of resources, and strategies to facilitate classroom enactment. To examine the contrast in professional development experience between treatment and control, we relied on the teacher survey.

Findings / Results:

The CRWP program was implemented largely as intended. In 20 of the 22 districts, at least 80% of ELA teachers participated in 45 or more hours of CRWP professional development each year. All 22 districts met implementation thresholds for the content of the professional development (i.e., the focus on argument writing). In terms of professional development strategies, all 22 districts used the NWP's formative assessment tool designed to evaluate students' source-based argument writing, while 18 of the 22 districts (82%) met the threshold for professional development events that focus on classroom enactment.

There were large, statistically significant differences between the treatment and control group in both the quantity and content of the professional development they received (Table 3). To better understand the distinction between the treatment and control

[2] The AWC has demonstrated high inter-rater reliability (overall 90% agreement across attributes), test-retest reliability, and internal consistency (Cronbach's $\alpha = .97$) (Bang, H. J., 2013). The final paper will include inter-rater reliability data from this scoring using the AWC-SBA.

condition, we asked teachers about their professional development experiences during year 2. Treatment teachers reported receiving nearly ten times as many hours of writing professional development as control teachers (63 hours compared to 6.4 hours) and reported a wider variety of experiences in line with both CCSS argument writing standards and best practice literature on teacher professional development.

Treatment teachers spent slightly more instructional time spent on writing, though they were substantially more likely than control teachers to focus that time on argument writing (Tables 4 and 5). The time teachers spent on writing differed very little, with treatment teachers asking students to write somewhat more frequently and over similar periods of time (about 30 minutes, when writing took place) as control teachers. The work students were asked to do differed more dramatically, with treatment teachers reporting a greater focus than control teachers on developing a claim, evaluating evidence that could be used in support of this claim, developing an argument in support of this claim, and practice writing the argument.

CRWP had positive and robust impacts on student argument writing (Table 6). We estimate an impact of .22 on the content attribute of the AWC-SBA, significant at the $p < .005$ level and robust to alternative model specifications (approximately a .20 effect size).[3] The impact estimate on the structure attribute is similar in size and significance. The point estimates for stance and conventions are qualitatively smaller, though not statistically different from those estimated on content and structure.

Conclusions:

CRWP impacted student outcomes on a particularly complex task—writing an argument supported by reasoning and developed through the use of evidence from source material. This type of argument writing has been identified as critical to college and career readiness and is central to the Common Core standards for English language arts and literacy.

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

This evaluation 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 are suggestive for developing theory about high quality professional development. This study adds rigorous experimental evidence to a body of literature that indicates that professional development that focuses on more than one aspect of instructional capacity—in this case, CRWP attended to both teachers and instructional materials—is more likely to lead to meaningful student learning. As such, the findings we will present have important implications for practitioners, policymakers, and program developers.

[3] We established student baseline equivalence using the standardized baseline AWC-SBA content scores and an identical structural model (point estimate=.06, p >.2).

Appendices

Appendix A. References

Bang, H. J. (2103). Reliability of National Writing Project's Analytic Writing Continuum Assessment System. *The Journal of Writing Assessment 6* (1).

Calkins, L., Ehrenworth, M., & Lehman, C. (2012). *Pathways to the common core: Accelerating achievement*. (1 ed.). Portsmouth, NH: Heinemann.

College Board. (2004). *Writing: A ticket to work...Or a ticket out. A survey of business leaders*. Report of The National Commission on Writing for America's Families, Schools, and Colleges.

Cohen, D. K., & D. L. Ball (1999). *Instruction, Capacity, and Improvement.* CPRE Research Report Series RR-43. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.

Cutler, L., & Graham, S. (2008). Primary grade writing instruction: A national survey. *Journal of Educational Psychology*, *100,* 907–919.

Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2011). Middle School Mathematics Professional Development Impact Study: Findings After the Second Year of Implementation (NCEE 2011-4024). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, Michael S., Stephanie Cronen, Marian Eaton, Anja Kurki, Meredith Ludwig, Wehmah Jones, Kazuaki Uekawa, Audrey Falk, Howard Bloom, Fred Doolittle, Pei Zhu, & Laura Sztejnberg (2008). The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement (NCEE 2008-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

National Center for Education Statistics. (2012). *The nation's report card: Writing 2011* (NCES 2012–470). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Persky, H., Daane, M., & Jin Y. (2003). *The Nation's Report Card: Writing 2002*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Rothman, R. (2011). *Something in common: the Common Core Standards and the next chapter in American education*. Cambridge, MA: Harvard Education Press.

Yoon, K., Duncan, T., Lee, S., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement. *Issues and Answers Report*, REL 2007 – No. 033. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Appendix B. Tables and Figures

Table 1: District Enrollment Sizes and Descriptive Statistics, by Treatment Status

	Treatment	*Control*
Limited English Proficency		
Mean	3%	2%
Minimum	0%	0%
Maximum	22%	18%
s.d.	6%	4%
n	22	22
Free/ Reduced Price Lunch Eligibility		
Mean	65%	67%
Minimum	46%	37%
Maximum	94%	99%
s.d.	14%	16%
n	21	22
Students of Color		
Mean	35%	39%
Minimum	5%	2%
Maximum	98%	98%
s.d.	29%	34%
n	22	22
Total District Size (K-12)		
Mean	2,569	2,195
Minimum	95	119
Maximum	15,919	6,618
s.d.	3,282	1,832
n	22	22

Source: National Center for Educational Statistics Common Core of Data, 2010-11

Table 2: Outcome Descriptives, by Treatment Status and Administration Time

	Baseline, Unstandardized		Baseline, Standardized		Outcome, Unstandardized	
	Treatment	*Control*	*Treatment*	*Control*	*Treatment*	*Control*
Content						
Mean	2.6	2.6	0.0	0.0	3.0	2.8
Minimum	1.0	1.0	-2.0	-2.0	1.0	1.0
Maximum	6.0	6.0	3.4	3.4	6.0	6.0
s.d.	1.0	1.0	1.0	1.0	1.1	1.1
n	1,259	1,227	1,259	1,227	1,259	1,227
Structure						
Mean	2.5	2.5	0.0	0.0	3.0	2.7
Minimum	1.0	1.0	-1.9	-1.9	1.0	1.0
Maximum	6.0	5.5	3.6	3.2	6.0	6.0
s.d.	1.0	1.0	1.0	1.0	1.1	1.1
n	1,259	1,227	1,259	1,227	1,259	1,227
Stance						
Mean	2.7	2.6	0.0	0.0	3.1	2.9
Minimum	1.0	1.0	-1.9	-1.9	1.0	1.0
Maximum	6.0	6.0	3.1	3.3	6.0	6.0
s.d.	1.1	1.1	1.0	1.0	1.2	1.1
n	1,259	1,227	1,259	1,227	1,259	1,227
Conventions						
Mean	2.7	2.8	0.0	0.0	3.2	3.1
Minimum	1.0	1.0	-1.9	-1.9	1.0	1.0
Maximum	6.0	6.0	3.0	3.3	6.0	6.0
s.d.	1.1	1.1	1.0	1.0	1.2	1.2
n	1,259	1,227	1,259	1,227	1,259	1,227

Table 3: Treatment-control Contrast

Survey question	Treatment	Control		*p* - value	Block *n*	District *n*	Teacher *n*
Since the end of last school year, how many hours of PD on writing instruction have you received?	63.41	6.41	***	0.00	22	44	301
Received PD that provided lessons… for teaching focused on argument writing (1=Yes)	98%	47%	***	0.00	22	42	214
Received PD focused on writing from source material (1=Yes)	91%	50%	***	0.00	22	42	214
To what extent did the PD involve… [a]							
Participating in a lesson, activity, or strategy as a learner	2.82	2.21	***	0.00	22	42	213
Engaging in writing myself	2.65	1.83	***	0.00	22	42	212
Reading and/or discussing professional literature	2.23	1.93	**	0.01	22	42	214
Analyzing student work	2.86	2.07	***	0.00	22	42	212
Planning how to implement what was learned in PD in my classroom	2.78	2.50	***	0.00	22	42	213
Designing tasks / assignments	2.54	2.25	**	0.00	22	42	214
Observing a demonstration lesson in a classroom with students	2.11	1.60	***	0.00	22	42	212
Co-teaching or collaborating with an instructional coach	2.26	1.47	***	0.00	22	42	212
Integrating or adapting new approaches to writing instruction with existing materials	2.56	2.10	***	0.00	22	42	214
Sharing my expertise with other teachers	2.36	1.79	***	0.00	22	42	213
To what extent do you dis(agree): I received adequate professional development to teach writing [b]	4.27	2.97	***	0.00	22	44	305

[a] Scale: 1=Not at all, 2=Minor exent, 3=Major extent

[b] Scale: 1=Strongly disagree, 2=Disagree, 3=Neither agree nor disagree, 4=Agree, 5=Strongly agree

Table 4: Teacher Outcomes, Daily Instructional Log

Log Question	Treatment	Control		*p* - value	Block *n*	District *n*	Teacher *n*	Log *n*
Did you ask students to write? (1=Yes)	94%	89%	*	0.03	22	44	341	2,938
How long did they write (in minutes)?	32.66	31.11		0.46	22	44	334	2,382
How much did they write?[a]	2.80	2.50	**	0.01	22	44	332	2,305
Student work included:								
Practice writing an argument	26%	7%	***	0.00	22	44	334	2,463
Generate an argument in support of a claim	26%	10%	***	0.00	22	44	334	2,463
Identify a claim and evidence in text	27%	26%		0.75	22	44	334	2,463
Develop a claim	36%	17%	***	0.00	22	44	334	2,463
Evaluate the credibility of evidence	19%	10%	***	0.00	22	44	334	2,463
Support a claim with evidence from a text	44%	41%		0.53	22	44	334	2,463
Support a claim with evidence from personal experience	13%	6%	**	0.01	22	44	334	2,463
Elaborate upon evidence used to support a claim	24%	15%	**	0.00	22	44	334	2,463

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

Table 5: Teacher Outcomes, Survey

Survey Question: *When writing, how much emphasis did you place on…* [a]	Treatment	Control		*p* - value	Block *n*	District *n*	Teacher *n*
Using language effectively	3.90	3.96		0.56	22	44	308
Conventions and usage	3.59	3.97	***	0.00	22	44	308
Using words, phrases, and clauses to link the major sect	3.66	3.66		1.00	22	44	308
Using a style and tone appropriate for the audience and purpose	3.86	3.77		0.43	22	44	307
Writing introductions and conclusions	4.18	4.14		0.73	22	44	309
Organizing ideas	4.43	4.31		0.23	22	44	308
Developing a claim	4.71	4.04	***	0.00	22	44	309
Connecting evidence to a claim	4.76	4.19	***	0.00	22	44	309
Selecting evidence from source material	4.63	4.15	***	0.00	22	44	308
Introducing and commenting on quoted text or source material	4.49	3.68	***	0.00	22	44	309
On-demand writing in response to text	4.04	3.67	**	0.00	22	44	308

[a] Scale: 1=No or almost no emphasis, 2=Minor emphasis, 3=Some emphasis, 4=Significant emphasis, 5=Heavy emphasis

Table 6: Student Impact Estimates

	Impact Estimate		s.e.	*p* - value	Block *n*	District *n*	Student *n*
Impact on each attribute:							
Content	0.22	**	(0.08)	0.01	22	44	2,486
Structure	0.22	**	(0.07)	0.00	22	44	2,486
Stance	0.17	*	(0.08)	0.03	22	44	2,486
Conventions	0.15	*	(0.07)	0.05	22	44	2,486
Alternate model specifications (outcome = content)							
Fixed block effects, control for student ahcievement only	0.21	***	(0.06)	0.00		44	2,486
Random block effects, control for student and district prior achievement	0.19	**	(0.07)	0.01	22	44	2,486
Random block effects, control for district prior achievement only	0.19	**	(0.07)	0.01	22	44	2,486
Random block effects, no prior achievement controls	0.24	**	(0.09)	0.01	22	44	2,486