Abstract Title Page

Title: Impact Evaluation of National Writing Project Professional Development Program

Authors and Affiliations: H. Alix Gallagher, Katrina Woodworth, Teresa McCaffrey, Christina J. Park, and Haiwen Wang (SRI International)

Abstract Body

Problem

Improving teacher effectiveness is a key strategy to ensure student readiness for college and careers and to address achievement gaps and persistent low performance. In response to the new Common Core State Standards for English Language Arts (CCSS-ELA) the National Writing Project (NWP) created a professional development (PD) program to support third-, fourth-, and fifth-grade teachers with their writing instruction. In partnership with the NWP, researchers sought to study the implementation of the program, estimate its effects, understand how context affected implementation, and provide formative feedback about the program.

Research Questions

The study had four primary research questions:

1. Was the PD program implemented with fidelity?
2. What impact did the PD program have on teacher practice?
3. What impact did the PD program have on student writing?
4. How did contextual factors influence teachers' uptake of new ideas in their classrooms?

Intervention

The NWP is a network of approximately 200, university-based Local Writing Project (LWP) sites that deliver PD throughout the country. LWP sites share a common model that includes university faculty working in collaboration with K–12 expert teachers.

The NWP received a Supporting Effective Educator Development (SEED) grant in 2012. With that grant, they designed a PD program to improve writing instruction in high-need elementary schools in line with the CCSS-ELA. The program model required that:

- 75% of third-, fourth-, and fifth-grade teachers in program schools participate in 40 hours of writing PD during the 2012–13 school year.
- The PD focus on the CCSS-ELA, most specifically in supporting teachers in teaching argument writing, which is one of the major shifts the CCSS-ELA require of teachers of writing.
- The PD use multiple delivery modes, in line with consensus in the field about features of effective PD (e.g., Yoon et al., 2007).

Setting

The NWP issued an RFP to LWP sites interested in participating in the SEED program and evaluation and actively recruited sites that had a track record of providing intensive in-service PD with elementary teachers and experience in high-need schools. NWP launched the program in partnership with 14 LWP sites, located in 13 states, including 7 in the South, 3 in the Midwest, 2 in the West.

Research Design

The evaluation was designed as a cluster randomized controlled trial in which schools were assigned to either the treatment or control condition. LWP sites were asked to recruit *pairs* of similar schools based on student achievement, student demographics, district (when feasible),

and other non-observable factors (e.g., reputation). All schools agreed to be randomly assigned to either participate in the SEED program during 2012–13 or receive a $3000 stipend in exchange for delaying participation in PD focused on writing or the CCSS-ELA standards for the 2012–13 school year, except for PD required by the district or the state. Additionally, control schools were eligible to participate in the SEED program during the 2013–14 school year. Once all schools were recruited, they were blocked by pair and randomized.

The evaluation drew on multiple sources of data, including detailed information on the nature and frequency of, and participation in, the SEED PD, observations of PD, interviews with the full range of program participants and PD providers, a teacher survey, and student writing in response to an on-demand prompt. Outcome measures—the teacher survey and student writing prompt—were administered in the fall (baseline) and spring (impact after 1 year).

Study Participants

LWP sites identified at least one *pair* of high-poverty schools, defined as a school in which at least 50% of students are from low-income families. The 14 LWP sites recruited 22 pairs of elementary schools for the study. During the winter of 2013, one school attrited; researchers dropped its matched pair, resulting in a final sample of 13 LWP sites and 42 schools (21 pairs). The SEED program targeted all third-, fourth-, and fifth-grade teachers in program schools.

Program and control schools had similar total enrollment, percent of students eligible for free or reduced-priced lunches, and percent of students classified by their states as English language learners (insert Table 1 here). On average 75% of students in schools qualified for free or reduced-priced lunches. In addition, the mean percentage of students scoring proficient or above appears similar across program and control schools, across all grades (insert Table 2 here). Finally, on average, teachers in the program and control schools had comparable years of teaching experience (insert Table 3 here).

Data Collection and Analysis

To measure program impacts on students, researchers administered on-demand writing prompts to all students in mainstream third-, fourth-, and fifth-grade classes in both program and control schools in the fall of 2012 and spring of 2013. Prior to scoring student writing in response to prompts, researchers drew a sample of 10 students per teacher who had completed fall and spring prompts. Excluding students who left or joined a class between the fall and spring prompt administrations limits the generalizability of findings. However, it does not bias the findings because student attrition was likely exogenous to treatment status due to school-level randomization, the fact that the study lasted 1 school year, and the fact teacher PD rarely causes student mobility. Analyses of the final analytic samples (both teacher and student) showed baseline equivalence between the groups on demographic and outcome measures.

Student responses were scored by expert upper elementary writing teachers using the NWP's Analytic Writing Continuum (AWC), which defines six levels of performance on a holistic measure and six attributes: Content, Structure, Stance, Sentence Fluency, Diction, and Conventions. The study's confirmatory analysis used an average of these six attributes. Prior research has shown that the AWC has good inter-rater reliability and has been reliable across scoring events (Bang, 2013; National Writing Project, n.d.; Swain & LeMahieu, 2012). The researchers applied hierarchical modeling (in SAS) to test whether assignment to the program condition impacted student scores on the writing prompts. For each outcome indicator, a two-level hierarchical linear model with student and school levels was posited, with the outcome at

the student level and program identification as a predictor at the school level. Researchers made adjustments for the baseline outcome and student grade level. The model also included site indicators to control for site effect. The coefficient of the program identification indicates the effect of program on the student writing outcome. The general model is:

$$Outcome = b_0 + b_1 Partnership + b_2 Baseline\ Outcome + b_l lth\ Grade\ Level + b_k kth\ Site\ Indicator + e + r$$

The terms *e* and *r* are random error terms at the student and school levels, respectively. The research team administered an online teacher survey to all third-, fourth-, and fifth-grade teachers in the treatment and control schools during the fall of 2012 and spring of 2013. The survey focused on teachers' experiences in writing PD, teachers' beliefs about writing instruction, teacher instructional practices, and teacher characteristics. Researchers applied hierarchical modeling to test whether assignment to the program condition impacted teacher outcomes. For each outcome indicator, a two-level hierarchical model with teacher and school levels was posited, with the outcome at the teacher level and program identification as a predictor at the school level. Researchers adjusted for the baseline outcome and baseline school average of the outcome indicator. Researchers also included site indicators in the model to control for site effects. The coefficient of the program identification indicates the effect of program on the teacher outcome. The general model is:

$$Outcome = b_0 + b_1 Partnership + b_2 Baseline\ Outcome + b_2 \overline{Baseline\ Outcome} + b_k kth\ Site\ Indicator + e + r$$

The terms *e* and *r* are random error terms at the teacher and school levels, respectively.

To understand program implementation and context, researchers collected data from LWP sites in September 2012, January 2013, and June 2013 on teacher attendance at PD events during the preceding 'semester.' LWP sites also provided information on the duration, content focus, and features of PD events. Additionally, researchers conducted phone interviews in the fall with a small sample of teachers and administrators in program and control schools, followed by spring phone interviews in control schools, and more intensive site visits (including interviews with a broader sample of teachers and administrators and observations of PD) at program schools.

Findings

LWP sites implemented the SEED program with fidelity in terms of the hours of PD provided to treatment teachers (insert Figure 1 here). Additionally, the SEED program provided teachers with more PD in writing and PD more focused on CCSS-ELA and research-based writing practices than the PD experienced by teachers in control schools (insert Tables 4-6 here).

The PD had positive impacts on some teaching practices (e.g., time students spent writing, frequency teachers taught strategies for effective argument writing, frequency that teachers used writing to help students learn other subjects) but did not impact other teaching practices (e.g., time spent teaching writing, length of writing students were asked to do, frequency with which students were asked to edit or revise writing) (insert Tables 7-11 here).

There was no evidence of impact on the quality of student writing (insert Table 12 here). Interview data suggest that teachers' nascent understanding of the type of writing instruction implied by the CCSS-ELA was one inhibitor to the type of more substantial shifts in writing

instruction that might lead to improved student outcomes. Additionally, limited time for writing instruction due to existing norms and practices and state tests that either did not focus on ELA or assessed student skills through short-answer or multiple choice items also limited teachers' willingness to devote more time to writing instruction.

Conclusions

Despite the research-based attributes of the SEED PD (see Yoon et al., 2007) and its influences on program school teachers' instructional practices, SEED PD did not impact student argument writing as measured by on-demand prompts scored on the National Writing Project's AWC. The lack of evidence of impact on student outcomes raises the question: why not? Looking across our data on teacher practices, what we found stands in contrast to what the writing instruction research (i.e., Graham et al., 2012) suggests students should have. At the most basic level, students did not have access to the opportunities to learn to write that research indicates they need. Interview data suggest that limited time devoted to writing is partially a legacy of NCLB (and related state and local instructional policies) and the tendency of schools to narrow the curriculum to focus on the most heavily-tested subjects. While the CCSS-ELA send different signals about writing instruction, implementation was in a very early stage and assessments were not yet aligned with the new standards. It seems likely that absent a surrounding context that is highly supportive of teacher learning and change, 1 year of PD cannot sufficiently alter instructional practices enough to impact student outcomes.

Based on the evaluation's key findings and an understanding of the context in which the National Writing Project's SEED program was implemented, this paper offers implications for state and district policymakers and school leaders, PD providers, and researchers.

State and district policymakers and school leaders. Research suggests that teachers' decisions (e.g., to change their instructional practices in writing) are influenced by the broader system of ideas, incentives, and sanctions present in their instructional context (Smith & O'Day, 1991). This research implies that if policymakers hope to see students moving towards the CCSS-ELA in writing, they will need to change the instructional policies and expectations that currently prevent writing from taking a more prominent role in instruction.

Professional development providers. Research on cognition and how it affects teachers' responses to substantially different ideas about instructional practices frames implications for PD providers. Spillane, Reiser & Reimer (2002) reviewed studies of past attempts to use standards to reform teaching to more inquiry-oriented approaches. Spillane et al.'s work suggests that a challenge for PD providers in the early stages of this reform may be teachers' level of understanding of what instruction aligned with new standards would actually look like. PD providers will need to help teachers envision the destination (i.e., a research-based instructional environment for student writing) as well as the path for moving from their current practices to those that are in line with the CCSS-ELA.

Researchers. Our data clearly show that the impact of SEED PD cannot be understood absent data on the context in which it was implemented. Given the context-related constraints for change, it might have been impossible for 40 hours of teacher PD on writing instruction, on its own, to have measurably impacted student writing in 1 school year. The implication for researchers, including those implementing randomized controlled trials, is the need to collect data not only on implementation and impact but also on context. In our study, some of the most compelling data on context was qualitative data, which many researchers collect sparsely, if at all, when conducting randomized controlled trials.

Appendices

Appendix A. References

Bang, H. J. (2013). Reliability of National Writing Project's Analytic Writing Continuum Assessment System. *Journal of Writing Assessment, 6(*1*).*

Graham, S., Bollinger, A., Booth Olson, C., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers: A practice guide* (NCEE 2012- 4058). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch.

National Writing Project. (n.d.). *National Writing Project offers high-quality writing assessment services*. Analytic Writing Continuum Brief. Berkeley, CA: Author.

Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing, 1990 yearbook of the Politics of Education Association*. London and Washington, DC: Falmer Press, 233–267.

Spillane, J., Reiser, B., & Reimer, T. (2002). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *Teachers College Record, 98*(3), 387–431.

Swain, S. & LeMahieu, P. (2012). Assessment in a culture of inquiry: The story of National Writing Project's analytic writing continuum. In N. Elliot and L. Perelman (Eds). *Writing assessment in the 21st Century: Essays in honor of Edward White*. New York: Hampton 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. Retrieved from http://ies.ed.gov/ncee/edlabs.

Appendix B. Tables and Figures

Table 1
Mean school-level demographic characteristics (baseline)

	Program	Control
Total enrollment	472 (160) *n* = 21	530 (207) *n* = 21
Percent of students eligible for free or reduced-priced lunch	73% (0.18) *n* = 20	75% (0.15) *n* = 20
Percent of students classified as English language learners or limited English proficient	23% (0.15) *n* = 9	22% (0.12) *n* = 8

Notes: (1) The SEED grant required schools to serve at least 50% of students who were eligible for free or reduced-price lunches under the Richard B. Russell National School Lunch Act or from low-income families as determined by using one of the criteria specified under section 113(a)(5) of the Elementary and Secondary Education Act of 1965, as amended. (2) These calculations exclude certain schools for which data could not obtained from the state department of education websites. Enrollment data was available for all 42 schools. There was one program and one control school for which no data on the number of students eligible for free or reduced-priced lunch could be obtained. There were also 25 schools for which no information on the number of students classified as English language learners. (3) Standard errors in parentheses.

Source: Publicly available data accessed from state education department websites in February 2014.

Table 2
Mean school-level percentage of students scoring proficient or above on state standardized English language arts or reading tests (baseline)

	Program	Control
Grade 3	55% (0.21) *n* = 21	61% (0.19) *n* = 21
Grade 4	61% (0.20) *n* = 21	64% (0.20) *n* = 21
Grade 5	61% (0.17) *n* = 21	57% (0.21) *n* = 21

Note: Standard errors in parentheses.

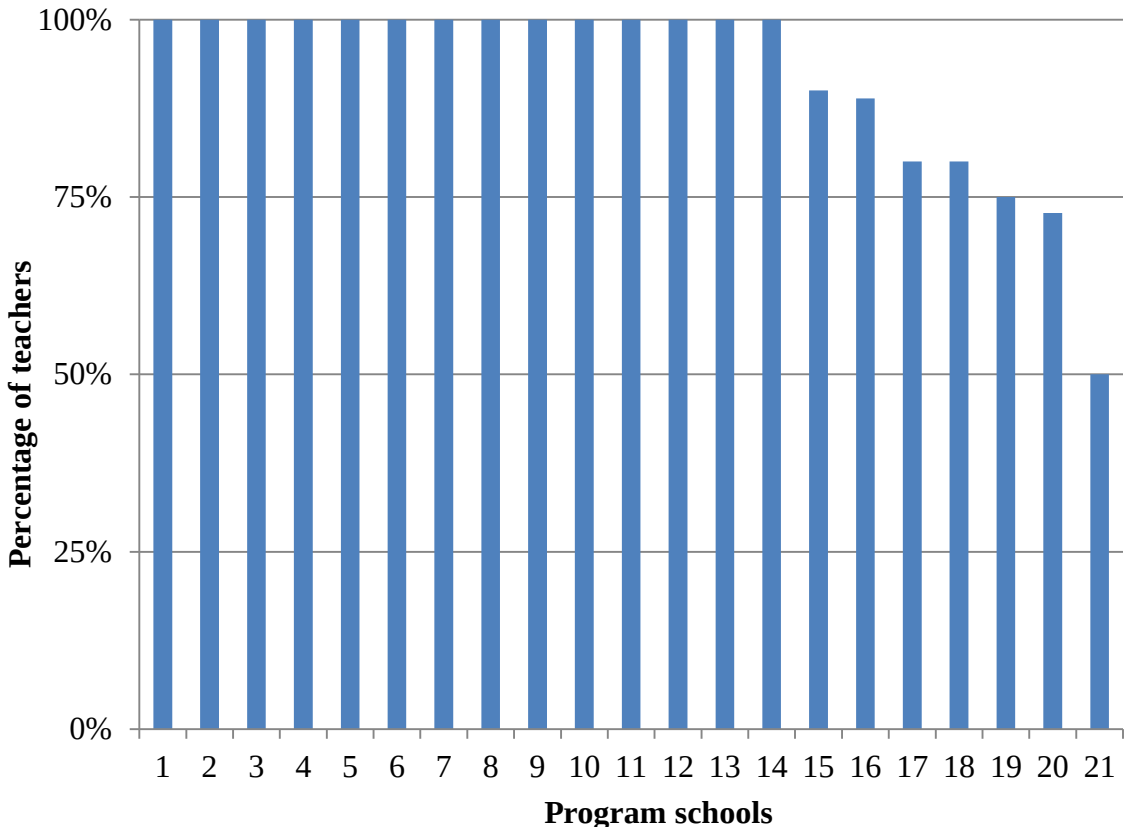
Source: Publicly available data accessed from state education department websites in February 2014. Assessments and proficiency thresholds are different in every state. Within each state the number of program and control schools was the same.

Table 3
Mean number of years spent teaching, including the current year (2012–13) (baseline)

	Program	Control
Years spent teaching	14.06 (0.97) *n* = 140	12.75 (1.19) *n* = 147

Note: Standard errors in parentheses
Source: Fall 2012 teacher survey.

Figure 1
Teachers' rates of participation in 40 or more hours of professional development, by SEED program school



Source: Professional development monitoring.

Table 4
Teachers' reports on the hours of writing professional development they participated in during the study year (2012-13) (including summer 2012)

	Program	Control	*p*-Value
Hours of writing professional development	61.13	8.32	<.01
	(18.26)	(1.67)	
	n = 138	*n* = 164	

Note: Standard errors are in parentheses.
Source: SEED teacher surveys.

Table 5
Frequency with which teachers identified a CCSS-ELA foci of professional development, among those who participated in professional development on the CCSS-ELA

	Program	Control	*p*-Value
An introductory overview describing the CCSS-ELA	83% 1.55 (0.28) *n* = 119	88% 1.95 (0.32) *n* = 88	.34
A crosswalk of my state's current standards and the CCSS-ELA (e.g., training highlighting similarities and differences between standards)	64% 0.58 (0.26) *n* = 119	61% 0.46 (0.22) *n* = 88	.68
Professional development on the shifts in instructional practice that are called for in the CCSS-ELA related to writing	52% 0.07 (0.24) *n* = 119	43% -0.29 (0.24) *n* = 88	.27
Professional development that provided lessons to teach opinion/argument writing	68% 0.77 (0.31) *n* = 119	9% -2.26 (0.48) *n* = 88	<.01
Professional development that provided lessons (or other specific guidance) to teach to the CCSS-ELA related to writing	55% 0.19 (0.25) *n* = 119	24% -1.18 (0.28) *n* = 88	<.01

Note: Standard errors are in parentheses. Only teachers who reported participating in at least one hour of professional development in writing responded to these questions. A total of 136 teachers in the program group and 92 teachers in the control group reported receiving at least one hour of writing professional development.

Source: SEED teacher surveys.

Table 6
Teachers' reports on the focus on using models and qualities of effective writing in professional development (means), among those who participated in writing professional development

	Program	Control	*p*-Value
Using models (e.g., mentor texts) in teaching writing	2.72 (0.05) *n* = 132	2.00 (0.10) *n* = 87	<.01
Qualities of effective writing (e.g., structuring effective sentences and paragraphs, using appropriate conventions, writing strong conclusions, word choice)	2.47 (0.07) *n* = 133	2.11 (0.10) *n* = 85	<.01

Note: Standard errors are in parentheses.Only teachers who reported participating in at least one hour of professional development in writing responded to these questions. A total of 136 teachers in the program group and 92 teachers in the control group reported receiving at least one hour of writing professional development.

Source: SEED teacher surveys.

Table 7
Teachers' reports on minutes students spent on writing per week (307 teachers in 42 schools)

Fixed Effects	Coefficient	*SE*
Program	21.68 *	9.93
Baseline score	0.18 **	0.05
Imputation flag for baseline score	16.76	11.20
School average for baseline score	0.21	0.15
Intercept	75.95 *	32.20
Random Effects	**Variance Component**	
School mean	223.35	
Teacher effect	5365.51	

Note: Coefficients for site indicators are omitted from the table.

**p < 0.05, **p < 0.01*

Source: SEED teacher surveys.

Table 8
Teachers' reports on the frequency of teaching aspects of effective argument writing (309 teachers in 42 schools)

Fixed Effects	Coefficient	*SE*
Program	0.90 **	0.17
Baseline score	0.47 **	0.06
Imputation flag for baseline score	1.75 **	0.24
School average for baseline score	-0.22	0.18
Intercept	2.06 **	0.68
Random Effects	**Variance Component**	
School mean	0.10	
Teacher effect	1.35	

Note: Coefficients for site indicators are omitted from the table.
**p < 0.05, **p < 0.01*

This measure is a scale averaged from the following items: Teach students organizational strategies for expressing an opinion or making a claim in writing, Teach students about words and phrases specific to expressing an opinion in writing (e.g., certain uses of "because," "for instance," or "specifically"), Teach students how to convey a clear and consistent point of view on a topic, Provide opportunities for students to practice expressing an opinion or claim through speaking (e.g., oral presentation), Use models (e.g., mentor texts) to teach students how to express a written opinion or make an argument, and Teach students how to support their opinions with reasons and evidence.

Source: SEED teacher surveys.

Table 9
Teachers' reports on minutes spent teaching writing per week (305 teachers in 42 schools)

Fixed Effects	Coefficient	*SE*
Program	0.74	7.32
Baseline score	0.17 **	0.04
Imputation flag for baseline score	0.48	7.67
School average for baseline score	0.41 **	0.13
Intercept	49.19 *	19.10
Random Effects	**Variance Component**	
School mean	148.81	
Teacher effect	2499.32	

Note: Coefficients for site indicators are omitted from the table.
**p < 0.05, **p < 0.01*
Source: SEED teacher surveys.

Table 10
Teachers' reports on how frequently students write multiple related paragraphs of 2 or more pages (309 teachers in 42 schools)

Fixed Effects	Coefficient	*SE*
Program	0.23	0.22
Baseline score	0.29 **	0.07
Imputation flag for baseline score	0.75 **	0.26
School average for baseline score	0.35	0.19
Intercept	1.23	0.74
Random Effects	**Variance Component**	
School mean	0.21	
Teacher effect	1.77	

Note: Coefficients for site indicators are omitted from the table.
**p < 0.05, **p < 0.01*
Source: SEED teacher surveys.

Table 11
Teachers' reports on the frequency students engaged in revising text (focused on meaning, ideas, and organization) (311 teachers in 42 schools)

Fixed Effects	Coefficient	*SE*
Program	0.25	0.16
Baseline score	0.54 **	0.07
Imputation flag for baseline score	2.30 **	0.33
School average for baseline score	0.23	0.13
Intercept	0.38	0.72
Random Effects	**Variance Component**	
School mean	0.03	
Teacher effect	1.61	

Note: Coefficients for site indicators are omitted from the table.
**p < 0.05, **p < 0.01*
Source: SEED teacher surveys.

Table 12
HLM model for writing prompts average score (3,835 writing samples in 42 schools)

Fixed Effects	Coefficient	*SE*
Program	0.00	0.05
Baseline average score	0.56 **	0.02
Grade 4	0.22 **	0.04
Grade 5	0.19 **	0.04
Intercept	0.00	0.06
Random Effects	**Variance Component**	
Student effect	0.93	
School mean	0.11	

Note: Coefficients for site indicators are omitted from the table.
**p < 0.05, **p < 0.01*
Source: SEED student writing prompts.