Struggling Readers in Urban High Schools:

Evaluating the Impact of Professional Development in Literacy

Marco A. Muñoz

Jefferson County Public Schools

Thomas R. Guskey

Georgetown College

Jennifer Aberli

Jefferson County Public Schools

Note: An earlier version of the paper was presented at the
Annual Meeting of the American Educational Research Association,
New York City, March 2008.

Abstract

High school teachers generally have little background or experience in meeting the needs of struggling readers and yet face classrooms filled with increasing numbers of students with inadequate reading skills. This paper describes the evaluation of a modified Ramp Up reading program, a district-wide professional development program designed to help high school teachers improve the reading skills of their students. Participants included 40 Grade 10 English teachers from 18 schools in a metropolitan school district. Guskey's five levels Model of Professional Development Evaluation served as framework for the evaluation. Results revealed that participating teachers liked the program and experienced significant gains in their knowledge of procedures for teaching reading skills. Discrepancies were noted between perceptions of teachers and district and school administrators with regard to factors related to organizational support. Despite these differences, teachers implementing the program saw statistically significant improvements in measures of student learning, particularly when implementation fidelity was observed. Implications for professional development policy, practice, and evaluation are discussed.

Struggling Readers in Urban High Schools:

Evaluating the Impact of Professional Development in Literacy

Marco A. Muñoz, Thomas R. Guskey, and Jennifer Aberli

The objective of this investigation was to evaluate the effectiveness of the modified Ramp-Up Program, a district-wide professional development effort designed to help high school teachers improve the reading skills of their students. Ramp-Up is a two-year course that seeks to accelerate the learning progress of entering high school students who are two or more years behind grade level in English/Language Arts. The course assumes that students can decode text and are reading at least at a Grade 3 level. Activities focus on helping students make rapid progress toward becoming fluent readers, develop wider vocabularies, and comprehend grade level texts through a variety of instructional approaches: (a) independent reading (Allington, 2001; Beers, 2003); (b) read-aloud/think-aloud/talk aloud (Hahn, 2002; Richardson, 2000); (c) whole-group and small-group reading and writing instruction (Fountas & Pinnell, 1996; Pearson, 1994); and, (d) collaborative learning situations including cross-age tutoring and service learning (Labbo & Teale, 1990). Pilot testing showed that the course had a positive effect on high school students' scores on norm-referenced reading and language arts tests (Muñoz, 2007).

The effectiveness of this course as a professional development model was evaluated using the five levels outlined by Guskey (2000; 2001a; 2001b) for evaluating professional development activities in education. The first level assesses participants' reactions, while the second level focuses on participants' learning through pre- and post-measures of knowledge and skills specific to the program. The third level considers participants' perceptions of

organizational support to enable change, and the fourth level focuses on participants' use of new knowledge and skills at the classroom level. The fifth and final level assessed the impact of the professional development activities on student learning outcomes using results from statewide accountability assessments in reading.

Research on Reading at the Secondary Level

Since reading is arguably the most fundamental academic skill, research on reading has long been a prime focus of educational researchers. Modern educators have recognized reading competence as a vital skill in its own right and as a necessary skill to achieve competence in other academic subjects, including the humanities, science, mathematics, and the social sciences. The demands of an information-based economy have made reading central to the debate on school improvement programs, including the federal legislation governing Title I and NCLB.

The basic skill of reading is an important concern of primary school educators. However, reading performance is an issue at all educational levels, including middle and high schools. Educators and others have voiced concerns about the reading ability of secondary students, especially when low ability creates barriers to effective job performance and enrollment in post-secondary education (Peterson, Caverly, Nicholson, O'Neal, & Cusenbary, 2000).

Despite the concern about poor reading among many high school students, most improvement programs have focused reading instruction in the elementary grades. As Graves (1999) noted, "Reading for secondary students – in fact reading for students beyond the primary and lower elementary grades – gets relatively little attention" (p. 1). The federal, state, and local emphases on educational accountability are making educators more accountable for the outcomes of all students from primary school through high school. Poor secondary school

reading is a problem that needs to be addressed in light of the new demands placed on educators at all levels by the NCLB legislation. Among the promising approaches to improve secondary reading that Graves identified were "teaching for understanding" and the concept of scaffolding student learning (Graves & Graves, 1994).

The same approaches identified by Graves have been identified by others. Secondary reading programs must assist students to gain the technical skills to read effectively (e.g., decoding skill) and the broader intellectual skills needed to become effective self-motivated readers. As Shellard (2001) noted, the focus in secondary school becomes less "learning to read" than "reading to learn." In other words, students must gain reading proficiency in order to learn the other subjects in the curriculum.

If the aim is to help problem readers, what does an effective secondary school reading program look like? Peterson et al. (2000) provided a comprehensive resource guide to the secondary school reading educator seeking to create such programs and described four major themes in the literature on improving secondary reading, including (a) motivation, (b) decoding print, (c) comprehension, and (d) transactions with texts.

In particular, language comprehension is essential to both sustaining reading motivation and to understanding school subject-matter. Comprehension is aided by explicit teaching of strategies that assist comprehension. Such strategies include: (a) teaching students to connect what they know to the text by building mental models of what they read (Graesser, Millis, & Zwaan, 1997), and (b) teacher modeling of "think-aloud" (Davey, 1983) and having students practice such techniques so they can be self-regulated learners. One sign of proficient reading in secondary school is the ability of the student to engage in dialog with the author of the text. This usually requires explicit modeling by the teacher. If the student recognizes that the author of a

text is a person who is speaking to the reader, the student may engage in an ongoing set of transactions; for example, aesthetic appreciation of language and literary effects.

According to Peterson et al. (2001), effective secondary school reading programs would likely have several characteristics. These include extensive use of scaffolding – teacher assistance to allow students to construct meaning from what they read. Part of scaffolding is explicit instruction in strategies. For example, the teacher might engage in a "think aloud" about a reading passage and then guide students to do the same. Effective instruction involves frequent sustained practice with accompanying support and feedback. This means that teacher assessment of reading performance is needed to ensure student progress. Finally, students would have some choice of the materials they read, thus helping to sustain student interest and motivation. The Ramp Up program has incorporated the core principles of current secondary school research into their literacy program.

Ramp Up Theoretical Framework

The theoretical framework behind the Ramp Up Program stems from current research on high school literacy. The school district invested in the program in response to the needs of struggling readers entering the ninth grade and after piloting a small version of the program during the 2003-04 school year (Muñoz, 2007). The Ramp-Up Program includes the following elements:

Independent Reading. Each Ramp Up class opens with independent reading, in which students read a book of their own choosing at their ability level. Independent reading helps students to develop and improve reading fluency and gives them time to practice previously learned strategies (Allington, 2001; Beers, 2003).

Read-Aloud/Think-Aloud/Talk Aloud. Read-aloud/think-aloud/talk-aloud provides opportunities for students to hear proficient readers make explicit their thoughts and the problems they encounter as they read (Hahn, 2002; Richardson, 2000). Students observe excellent readers employing a variety of comprehension strategies: (a) activating background knowledge, (b) visualizing, (c) questioning, (d) determining importance, (e) making inference, (f) summarizing, and (g) monitoring for meaning. Upon completion of the Read-Aloud/Think-Aloud, students then participate in a Talk-Aloud in which they dialogue with the teacher and one another about what was read.

Work Period: Whole and Small Group Instruction. During whole group instruction, students learn a variety of reading and writing lessons. Examples include learning how to select a book, give a book talk, and use sticky notes as a tool for reading with purpose. Students then work in small groups with others who are at similar stages of reading development or who need to acquire similar strategies for reading success. The teacher engages with the small group with texts appropriate to their level. Students ask questions and make predications with the teacher as a guide so that they can successfully apply what they have learned when reading in other settings (Fountas & Pinnell, 1996).

Work Period: Writing Instruction. Students learn the stages of the writing process, including planning, drafting, editing, conferencing, and revising. They then write in the genres they are reading so they can learn to read *as writers* do. In this way, students learn to understand an author's choices so that they can then apply these choices in their own writing. The experience offers excellent support for students who struggle with reading. According to Pearson (1994), writing also provides excellent support for reading comprehension strategies and activities.

Cross-Age Tutoring. Pairing older students with elementary students for tutoring in reading provides a real-world opportunity for secondary students to practice literacy strategies and develop self-confidence as readers. As tutors, they have the chance to model the strategies they learned in Ramp Up. Their understanding and application of these strategies is enhanced and reinforced as they teach the strategies to younger students. Reading children's books (the oral performance) to the younger students is documented to increase the fluency and comprehension skills of the older reader (Labbo & Teale, 1990).

Evaluation's Theoretical Framework

The theoretical framework for the evaluation was Guskey's five level evaluation model of professional development outlined in *Evaluating Professional Development* (Guskey, 2000; 2001a; 2001b). These levels begin with participants' reactions to the experience (Level 1), consider participants' learning (Level 2), look at organization support and change (Level 3), document participants' use or implementation (Level 4), and finally consider the impact on student learning outcomes (Level 5).

Backward mapping techniques were used with Guskey's model of professional development, beginning with the desired endpoint and then working backward to determine what must be done to answer various evaluation questions (Muñoz, 2005). When planning professional development experiences and accompanying evaluation activities, Guskey (2001a; 2001b) stresses that the five levels of evaluation must be reversed.

Using these backward planning techniques, school leaders first considered the student learning outcomes they wanted to achieve (Level 5). In this case, they wanted most to improve students' reading comprehension. Then they sought to determine what instructional practices

and policies would most effectively and efficiently yield those outcomes (Level 4). Next, school leaders considered the aspects of organizational support that had to be in place for those practices and policies to be implemented (Level 3). After that they identified what knowledge and skills participants must have in order to implement the prescribed practices and policies (Level 2). Finally, leaders considered the best means by which participants might be helped to acquire that knowledge and skills (Level 1). What makes this process effective as a planning tool is that the decisions made at each level help shape those made at the next.

Method

This evaluation was primarily an exploratory, quantitative investigation. Supplementary qualitative techniques were used to clarify issues brought to light by the quantitative data. The data gathered included: Level 1 (pre- and post-satisfaction assessments), Level 2 (pre- and post-knowledge assessments), Level 3 (participant and administrator questionnaires), Level 4 (observation ratings), and Level 5 (student reading test scores).

Data from Level 1 (satisfaction assessments) were collected during the summer and at multiple times during the year for each professional development session teachers attended. The Level 2 assessments (pre-post assessments) were administered to all teachers who attended the summer Induction and Refresher professional development sessions. The Level 3 questionnaires were administered to instructional coaches, principals and teachers on one occasion in the fall. During the fall and spring observations, teachers' implementation (Level 4) was assessed using a rubric created to determine both the degree and quality of program implementation. The classroom observations were conducted by two trained instructional coaches. The nature of the observation rubric was a rating scale ranging from non-productive to fully-operational practice.

Impact at Level 5 was assessed using a matched treatment-control, pre-posttest design (Cook & Campbell, 1979; Rossi, Freeman, & Lipsey, 1999; Shadish, Cook, & Campbell, 2002). Students in high implementation classrooms were compared to students in low implementation classrooms; in addition, high and low implementation groups were compared to a comparison group from the same schools. Data at the school level for both treatment and control classrooms were analyzed using descriptive and inferential statistics. Analysis of Covariance (ANCOVA) techniques were used to control for prior achievement, with the treatment condition as the between-subjects factor, the Predictive Assessment Series (PAS) "pretest" scores as the covariate, and tenth-grade Kentucky Core Content Test (KCCT) Reading test scores as the dependent variable. Examining program effects on these criterion-referenced assessment scores was the principal focus of the evaluation. All data were analyzed using the Statistical Package for the Social Sciences (SPSS).

Data Sources

The district that served as the site of this program is located in a large metropolitan area and has 150 schools serving approximately 97,000 students. The district educates a high percentage of at-risk urban students with high poverty levels (i.e., over 55% subsidized meals and single-parent homes). The district has a student assignment plan based on managed choice, which facilitates the racial desegregation of its schools by providing students with transportation from their home neighborhoods to other parts of the district. In the state were the district is located, School-Based Decision Making (SBDM) is part of the educational reform effort. The SBDM have the responsibility to set school policy consistent with district board policy to provide an environment to enhance students' achievement. SBDM plays a significant role in the

selection of comprehensive school reform programs, such as Ramp Up. In this context, district officials can suggest academic programs and interventions, but the individual schools have control over the adoption of curricula and programs. Data sources included:

Level 1 – Participants' Reactions. All teachers who participated in district level professional development were required to complete an anonymous on-line evaluation using the district's professional development tracking system. The participation in the Ramp Up professional development was open to all high school English teachers in the district. Since no systematic selection procedure was established, the sample of participating teachers could not be characterized as random. The types of questions included in this evaluation measured teachers' satisfaction with the experience. This form consisted of 21 items that inquired about various aspects of the professional development experience: (a) content, (b) context, (c) process, and (d) results. Pilot testing showed it to have an overall reliability of .98.

Level 2 – Participants' Learning. All teachers who attended the summer induction and refresher professional development sessions were administered a questionnaire. The intent was to determine whether or not participants acquired the intended knowledge and skills from the professional development. These questions measured pre-and post-test knowledge and skills of participants. This form consisted of six Likert-type scale items, developed by the program coordinators with assistance of the Research Department. Pilot testing showed it to have a reliability of .95.

Level 3 – Organizational Support and Change. The evaluators attempted to measure the organization's support by using questionnaires. This evaluation instrument was given to Ramp Up teachers as well as to instructional coaches and principals/administrators to check for triangulation of data. The instrument included three Likert-type subscales: (a) four items related

to professional development (alpha = 76); (b) five items related to program implementation (alpha = .88); and, (c) six items associated to organizational support and change (alpha = .84). The reliability of all 15 items was .91.

Level 4 – Participants' Use of New Knowledge and Skills. Direct observations conducted by two trained observers measured the quality of program implementation and attempted to determine whether or not teachers effectively applied the new knowledge and skills acquired from the professional development. A rubric was designed based on the program components (1 = non-productive practice, 2 = limited practice, 3 = partially-operational practice, and 4 = fully-operational practice). Three scales are associated with the observation rubric: (a) two items related to academic standards (alpha = .69), (b) three items related to rituals and routines (alpha = .79), and (c) five items related to pedagogy, literacy, and assessment (alpha = .76). The overall reliability of the ten items combined is .88.

Level 5 – Student Learning Outcomes. Student learning outcomes based on the statewide reading test were used to measure whether or not student achievement were affected by the professional development provided to teachers. The achievement measure employed was the statewide test in Reading. The "diagnostic" test, used for matching treatment and control students on prior achievement, was the Predictive Assessment Series (PAS). The analyses compared high and low implementation classrooms as identified by the observation data (i.e., level 4). All tests were group-administered and scored following standardized procedures.

Results

The findings from this evaluation reveal generally positive effects from the Ramp Up professional development program at all levels of the evaluation.

Level 1 – Participants' Reactions. Level 1 data shows that teachers were very satisfied with the district-provided professional development. The scores on all evaluations were high indicating that teachers were pleased with the format and the information they received. A dependent-sample t-test (N = 40) indicated a significant difference between pre- (M = 2.75, SD = 1.08) and post-test (M = 4:35, SD = .53) results in participants' reactions to the professional development [t (39) = 13.60, p < .001)].

Level 2 – Participants' Learning. Data from this evaluation survey confirmed expected results. Teachers who taught Ramp Up earlier (during the pilot) and attended the refresher PD showed little difference in their before and after training scores. However, teachers new to the program who attended the 3 day induction Ramp Up experience showed a significant difference between pre- (M = 3.08, SD = 1.17) and post-test (M = 4:87, SD = .23) average composite scores in participants' learning [t (43) = 10.36, p < .001)].

Level 3 – Organization Support and Change. The organizational support and change questionnaires were given to instructional coaches (n = 12), principals/administrators (n = 29), and teachers (n = 40). When looking at the overall data on percent agreement, the instructional coaches and principals scored highest in all three areas: (a) professional development; (b) program implementation; and, (c) organizational support and change. When comparing instructional coaches and principals with Ramp Up teachers, it was noted that teachers scored the lowest on all three areas: (a) professional development (98%, 98%, and 92% respectively), (b) program implementation (95%, 97%, and 84% respectively), and (c) organizational support and change (97%, 99%, and 94%, respectively). All differences were tested, but only the area of program implementation showed an overall statistically significant difference among instructional coaches, principals/administrators, and teachers [F (2, 12) = 4.71, p = .03]; the

Bonferroni post hoc test indicated a mean difference between teachers (M = 83.63, SD = 11.35) and the instructional coaches (M = 95.28, SD = 6.48) and principals/administrators (M = 97.46, SD = 2.33). Results showed a gap between the district's instructional coaches and the schools' principals/administrators and teachers in the area of planning professional development (92%, 96%, and 75%, respectively); also, data indicated a gap in the area of program implementation in quality follow-up support from the district (100%, 95%, and 87% respectively) and schools (88%, 100%, and 76%, respectively) as well as in receiving appropriate resources needed (89%, 96%, and 68%, respectively).

Level 4 – Participants' Use of New Knowledge and Skills. The observations of teachers indicate there was an overall gain in program implementation from the fall observation to the spring. Teachers (n = 40) observed in the fall had overall lower implementation scores when compared to the spring observations (M = 2.78 and SD = .44 vs. M = 3.00 and SD = .46), indicating that follow-up support may have helped improve program implementation fidelity [$F(1,105) = 6.33$, $p < .05$]. The results of these observations indicated specific areas of program implementation where improvement needs to continue in the professional development provided to teachers. In the particular area of rituals and routines, it was observed statistically significant gains [$F(1, 105) = 9.35$, $p < .01$] from the fall to the spring observation (M = 2.92 and SD = .57 vs. M = 3.23 and SD = .46), including students demonstrating the ability to enter the classroom according to expectations of the program and the teacher adhering to the day in the course schedule.

Level 5 – Student Learning Outcomes. The level 5 student learning outcomes were studied in various ways. The first analysis was conducted to compare high vs. low implementation groups. The second analysis expanded the comparison by including an additional comparison group of students from the participating schools. The third and final

analysis involved evaluating the overall impact of the program on the high school system of the district under examination. Following is a description of these various analyses.

Part 1: High vs. Low Implementation

The sample included students in Ramp Up classrooms with high (i.e., treatment) and low (i.e., control) implementation. The sample was split based on the median results associated with the observation rubric. The participating teachers began implementing the Ramp Up program in the 2006-07 school year. Table 1 illustrates the key characteristics by which the treatment and control students were matched at the beginning of the school year (2006-07) for the evaluation of the impact of professional development activities on reading test scores. As shown, the matches were based on three variables, including previous diagnostic test scores in reading, race, and participation in the free or reduced-price lunch program (i.e., proxy for poverty). The comparability of the matches was evaluated using chi-square tests for categorical variables and ANOVAs for continuous variables (i.e., previous test scores). The comparability of the matches was evaluated using a 2 x 2 chi-square tests for categorical variables [poverty, χ (1, N = 214) = .17, $p > .10$; race, χ (1, N = 214) = 2.04, $p > .10$) and ANOVAs for continuous variables [previous test scores, F (1, 213) = 9.60, $p = .01$]. As result, only previous test scores were used as a covariate to adjust for initial differences between the treatment students and the comparison students.

Table 1

Treatment and Comparison Students Matched on Key Characteristics (N = 214)

	Treatment Students				Comparison Students			
	M	SD	n	%	M	SD	n	%
Previous Test	1048.3	13.4	149		1042.0	14.4	65	
Poverty								
Free/Reduced lunch			113	75.8			51	78.5
Paid lunch			36	24.2			14	21.5
Race								
Minority			83	55.7			43	66.2
Non-Minority			66	44.3			22	33.8

Note. Only students with complete demographic and testing data were included in the analysis. An aggregated matching procedure was utilized.

Level 5 Instrumentation

For this evaluation, criterion-referenced tests were analyzed. The criterion-referenced test was the KCCT in Reading (24 multiple choice, 6 open response items). The diagnostic test was the PAS (24 multiple choice items). The Predictive Assessment Series (PAS), a ThinkLink benchmark test, are considered reliable predictors of student performance on criterion referenced tests; ThinkLink PAS results can be used to gauge progress toward "Proficiency" levels as defined by each state under the requirements of NCLB (ThinkLink, 2007). KCCT Reading represents 14.5% of the 100% accountability formula used as part of the annual state assessment. KCCT in reading test was group-administered and scored following standardized procedures. Only those students who had complete data encompassing the PAS and KCCT Reading served as

the basis for the student- level matching component (Rossi, Freeman, & Lipsey, 1999). The PAS test was administered to ten graders at the beginning of the school year (2006-07). At the end of the 2006-07 school year, tenth graders took the KCCT Reading subtest.

Level 5 Findings

When conducting an analysis of covariance, the interactions between covariates and factors should be tested. Doing so will determine whether you have met the *homogeneity of regression slopes* assumption, which states that the regression slopes for all groups in your analysis are equal. This assumption is important because the means for each group are adjusted by averaging the slopes for each group so that group differences in the covariate are removed from the dependent variable. Thus, it is assumed that the relationship between the covariate and the dependent variable is the same at all levels of the independent variables. In order to meet the ANCOVA assumption, these interactions should not be significant; in our evaluation, we expected to see non-significant effects for the *group*PAS* interaction effects. The *group*PAS* effect had a small *F* statistic (1.20) and a large significance value (.28). Because this significance level is greater than .05, the homogeneity of regression assumption was met and we proceeded with the ANCOVA.

The ANCOVA performed on KCCT Reading yielded significance for the PAS covariate [$F(1, 213) = 65.37$, $p < .001$]. More importantly, a significant program effect [$F(1,213) = 4.23$, $p < .05$] was found between treatment and the comparison students. A caution that is necessary with the results associated with this comparison is the relatively small number of control students included in the analysis (Treatment = 149; Control = 65). Table 2 displays a summary of the KCCT Reading student achievement analysis. An effect size was computed for both unadjusted and adjusted means and standard deviations associated with the program participants.

Table 2

Unadjusted and Adjusted Means and Standard Deviations for Treatment and Matched-Control Comparisons on KCCT Scale Scores in 2006-07(N = 214)

Group	*N*	*M*	M_{adj}	*SD*	ES[a]	ES[b]
KCCT Reading						
Treatment	149	1035.5	1034.9	8.9	+0.49	+0.27
Control	65	1030.8	1032.2	11.0		

Note. [a]Effect sizes were computed from the unadjusted means; [b]Effect sizes were computed from the adjusted means.

Part 2: High vs. Low Implementation Compared to Control Group

The sample included students in the treatment (i.e., Ramp Up with high and low implementation) and a control group. The participating teachers began implementing the Ramp Up program in the 2006-07 school year. Table 3 illustrates the key characteristics by which the different treatment implementation levels and control students were matched at the beginning of the school year (2006-07) for the evaluation of the impact of professional development activities on reading test scores. As mentioned earlier, the low and high implementation group was selected based on the median split on the observation rubric. The control group, however, was selected from the same schools and from teachers working with students with similar characteristics; limitations in funding have not allowed implementation of Ramp Up in all classrooms of the participating schools. As shown, the matches were based on three variables, including previous diagnostic test scores in reading, race, and participation in the free or reduced-price lunch program (i.e., proxy for poverty). The comparability of the matches was evaluated using chi-square tests for categorical variables and ANOVAs for continuous variables

(i.e., previous test scores). The comparability of the matches was evaluated using 3 x 2 chi-square tests for categorical variables [poverty, χ (2, N = 340) = 6.16, $p > .05$; race, χ (2, N = 340) = 2.04, $p > .10$) and ANOVAs for continuous variables [previous test scores, F (2, 337) = 5.11, $p < .01$]. As a result, only previous test scores were used as a covariate to adjust for initial differences between the treatment students and the comparison students.

Table 3

Treatment and Comparison Students Matched on Key Characteristics (N = 340)

	High Implementation				Low Implementation				Comparison Students			
	M	SD	n	%	M	SD	n	%	M	SD	n	%
Previous Test	1048.3	13.4	149		1042.0	14.4	65		1047.6	13.8	126	
Poverty												
Free/Reduced			113	75.8			51	78.5			81	64.3
Paid lunch			36	24.2			14	21.5			45	35.7
Race												
Minority			83	55.7			43	66.2			74	58.7
Non-Minority			66	44.3			22	33.8			52	41.3

Note. Only students with complete demographic and testing data were included in the analysis. An aggregated matching procedure was utilized.

Level 5 Instrumentation

For this evaluation, criterion-referenced tests were analyzed. the criterion-referenced test was the KCCT in Reading (24 multiple choice, 6 open response items). The diagnostic test was the PAS (24 multiple choice items). The Predictive Assessment Series (PAS), a ThinkLink benchmark test, are considered reliable predictors of student performance on criterion referenced tests; ThinkLink PAS results can be used to gauge progress toward "Proficiency" levels as

defined by each state under the requirements of NCLB (ThinkLink, 2007). KCCT Reading represents 14.5% of the 100% accountability formula used as part of the annual state assessment. KCCT in reading test was group-administered and scored following standardized procedures. Only the students who had complete data encompassing the PAS and KCCT Reading served as the basis for the student- level matching component (Rossi, Freeman, & Lipsey, 1999). The PAS test was administered to tenth graders at the beginning of the school year (2006-07). At the end of the 2006-07 school year, tenth graders took the KCCT Reading subtest.

Level 5 Findings

When conducting an analysis of covariance, the testing the interactions between covariates and factors will determine whether you have met the *homogeneity of regression slopes* assumption, which states that the regression slopes for all groups in your analysis are equal. This assumption is important because the means for each group are adjusted by averaging the slopes for each group so that group differences in the covariate are removed from the dependent variable. Thus, it is assumed that the relationship between the covariate and the dependent variable is the same at all levels of the independent variables. In order to meet the ANCOVA assumption, these interactions should not be significant; in our evaluation, we expected to see non-significant effects for the *group*PAS* interaction effects. The *group*PAS* effect had a small *F* statistic (0.96) and a large significance value (.38). Because this significance level is greater than .05, the homogeneity of regression assumption was met and we proceeded with the ANCOVA.

The ANCOVA performed on KCCT Reading yielded significance for the PAS covariate [$F(1, 336) = 125.92$, $p < .001$]. More importantly, after adjusting for the strong effect of the covariate, a significant Program effect [$F(2,336) = 6.49$, $p < .01$] was found between treatment

and the comparison students. To assess pair-wise differences among the three levels for the main effect for group (i.e., high implementation/treatment, low implementation, control), the least significant difference (LSD) follow-up procedure was performed. The results indicated that the high implementation/treatment mean differed significantly from the low implementation/control mean (M Difference = 2.49, $p < .05$) and the control mean group (M Difference = 3.43, $p < .01$); the results indicated no significant difference between low implementation/treatment and control group (M Difference = .93, $p > .05$).

A caution that is necessary with the results associated with this comparison is the relatively small number of low Ramp Up implementation students (n = 65) included in the analysis when compared to the high implementation/treatment (n = 149) and control (n = 126) students. The smaller sample for the low implementation group also means lower levels of statistical power. Table 4 displays a summary of the KCCT Reading student achievement analysis. An effect size was computed for both unadjusted and adjusted means and standard deviations associated with the program participants.

Table 4

Unadjusted and Adjusted Means and Standard Deviations for Treatment and Matched-Control Comparisons on KCCT Scale Scores in 2006-07(N = 340)

Group	*N*	*M*	M_{adj}	*SD*	ES[a]	ES[b]
KCCT Reading						
High Implem.	149	1035.5	1035.0	8.9	+.49	+.26
Low Implem.	65	1030.8	1032.5	11.0		
Control	126	1031.8	1031.6	9.2	+.38	+.36

Note. [a]Effect sizes were computed from the unadjusted means; [b]Effect sizes were computed from the adjusted means.

Level 5 – School Trend Analysis on Student Learning Outcomes. Aggregated data on academic achievement in reading were calculated for high schools from several years prior to the Ramp Up implementation (2003-2004 school year) to the last round of assessment data available (2006-2007 school year). All data were abstracted from computerized files provided by the school district that served as the evaluation site. In this district, the state's performance levels are (a) Novice, (b) Apprentice, (c) Proficient, and (d) Distinguished; the reading academic index contributes 14.5% of the 100% accountability formula used as part of the annual state assessment. The statewide reading test was group-administered and scored following standardized procedures.

The following figures illustrate the program's effects. Figure 1 displays the percentage of Novice reduction while Figure 2 shows the percentage of Novice reduction associated with the racial gap in reading achievement. Figure 3 indicates the positive changes in the reading academic index since the implementation of Ramp Up in the district.

Figure 1

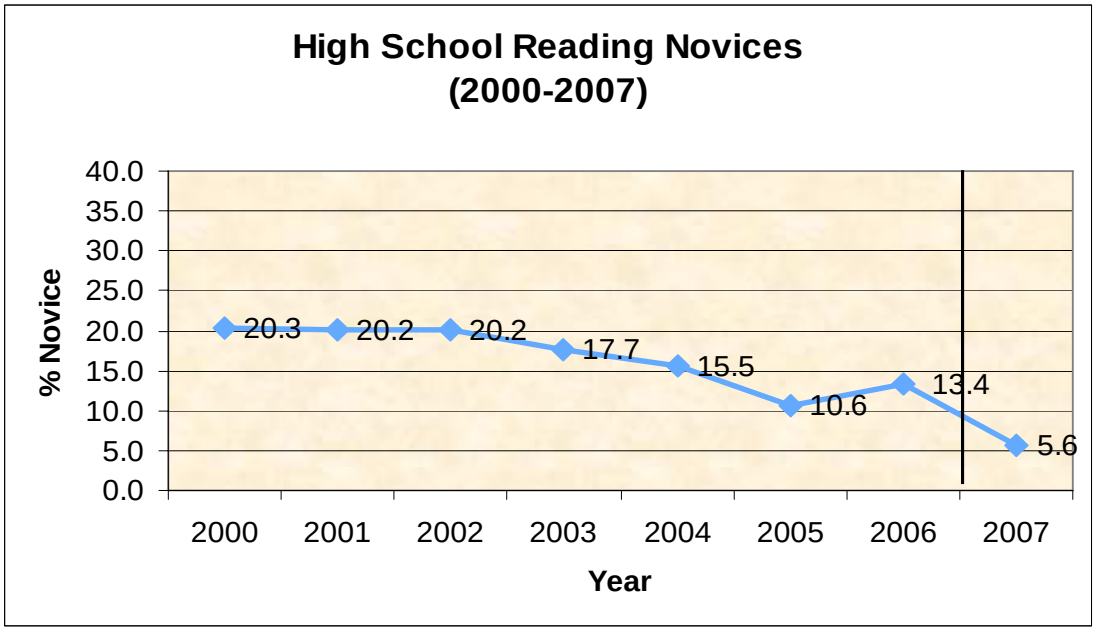


Figure 2

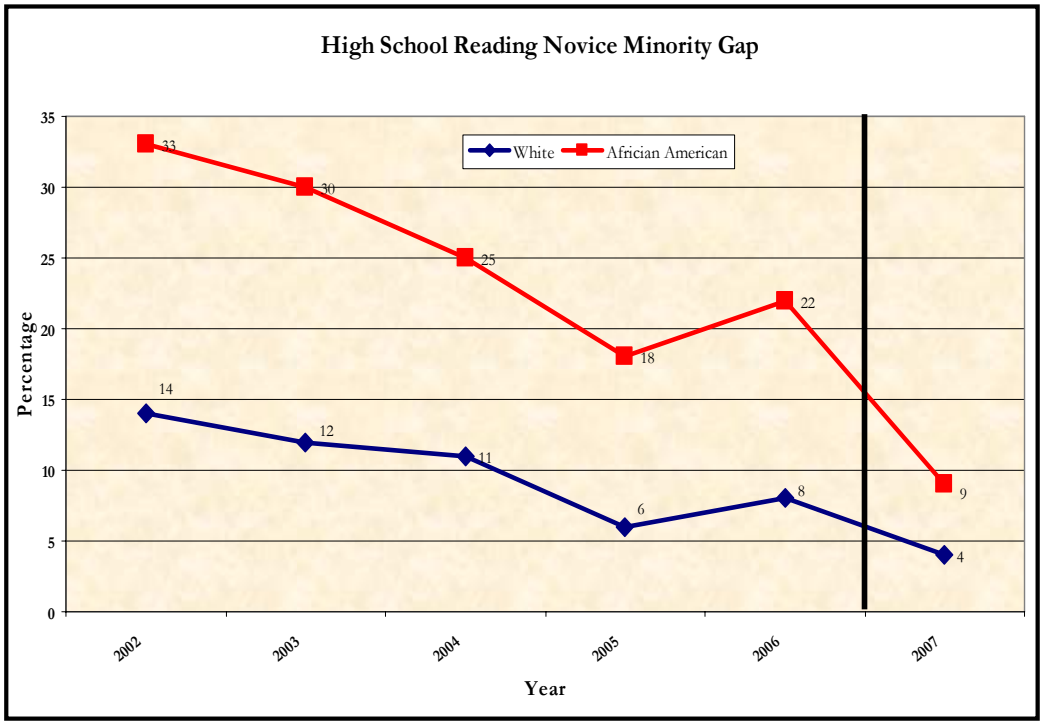
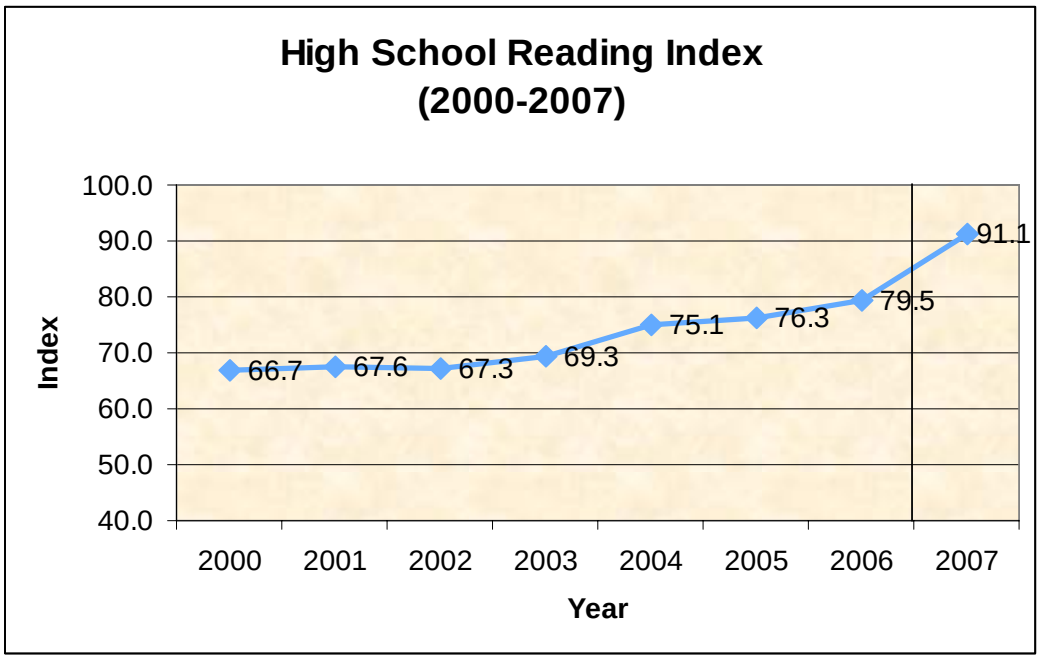


Figure 3



Discussion

Given the current high-stakes federal and state legislation, it is important to examine professional development programs and their impact on educational improvement and learning. The impact of professional development on student achievement is a critical element to facilitate better decision making from educational leaders associated with our public schools. Unfortunately, few evaluations analyzing the impact of professional development on student achievement are available to school leaders.

The concept of Ramp Up is to reform high school literacy instruction to ensure that every student learns to read at a high level of proficiency. Given the high importance of reading, data on the efficacy of literacy interventions at the high school level are important to report. The results of this evaluation showed favorable achievement outcomes for the students enrolled with teachers with high participation on the professional development activities associated with the Ramp Up program. The results were supportive of the professional development effects on academic achievement as measured by the state-mandated test scores in reading. On the KCCT Reading subtest in the tenth grade, Ramp Up participants were significantly higher than the matched control students. At the school level, it was noted a notable increase in the academic achievement in reading while a remarkable decrease in Novices (i.e., lowest level of performance in the statewide assessment).

Although many times professional development activities are judged based on supplementary individualized assessments, the ultimate judgment of the effectiveness of professional development programs has to be linked back to student-level measures of learning. Professional development activities that cannot be shown to have an impact on student learning outcomes and academic performance probably should be restructured or redefined.

Limitations and Implications for Future Work

As with any evaluation, the present preliminary findings are qualified by several factors. Since a randomized experimental design was not employed, firm causal conclusions cannot be reached. Sampling representativeness is limited by teacher choice to participate in the PD associated with Ramp Up. The program was introduced to the teachers as a way of addressing the needs of struggling readers in high schools. The teachers made their choice to participate based on their interest in learning about this particular approach that seeks to accelerate learning to read. As a result, it is difficult to isolate the impact of the PD participation from other possible explanations for student achievement changes. For example, the participation and outcomes of students might be affected by teachers' motivation to participate in the program.

Despite of confounding variables, the aggregated matching procedure and statistical controls should have established sufficiently strong internal validity for attributing differential group change more to the PD than to potentially confounding intervening factors (Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). Additionally, due to nature of the state assessment system and incomplete testing data, shrinkage of sample sizes for both the treatment and control groups was observed. Direct year-to-year comparisons of KCCT scores are difficult because the reading tests are given to particular accountability grades (e.g., only cohorts of grade 10 student take the reading test).

Subsequent study is needed to follow up these early, promising results with the Ramp Up professional development model. Given the dynamic school characteristics and internal factors affecting model implementation, further analysis using qualitative methods might clarify elements related to implementation quality and stakeholder buy-in. These are important elements in the discussion of any externally developed comprehensive reform model.

Conclusion

According to the *No Child Left Behind* (NCLB) legislation (U.S. Congress, 2001) there is a strong emphasis on accountability, defined in terms of student performance. An important implication of this NCLB legislation impacts the professional development provided by the district. No longer can we rely heavily upon the collection of PD Evaluation using Levels 1-4 to impact our PD, we must look more closely at how the PD experiences may lead to specific improvement in student performance.

Based on the gathering of data using Guskey's five levels of professional development evaluation (Guskey, 2000), significant effects were seen on some but not all measures. Still, the consistent direction of the effects showed benefits for Ramp Up teachers when compared to control teachers. As result, we are providing job-embedded PD to those teachers who continue to score low on the Level IV (observations) and Level V (student achievement) data. By providing this more tailored PD to the needs of these struggling teachers, we hope to better meet their instructional needs, thus improving student performance.

According to Guskey (2001a; 2001b), one way to provide better professional development is to plan "backward" by using his 5 level evaluation model. By planning backward, the first and foremost attention is focused on the ultimate goal of professional development: improved student learning outcomes. Secondly, by planning backward, the evaluation questions are considered before the PD begins, which leads to better design and implementation. If evaluators help school leaders begin their planning with what they want to achieve in terms of learning and learners, and then work backward from there, not only will planning be a lot more efficient, but evaluation efforts much more meaningful and easier to implement.

Finally, it will also be important to consider what other research has to say about effective professional development. We know that PD can not be a one-size-fits-all training with little follow-up support (Robb, 2002). Furthermore, high quality PD, that is sustained and content-focused, has a positive impact on student learning (Desimone et. al., 2006), which is our ultimate goal in PD. One of the anticipated challenges in moving toward a job-embedded PD model will be asking teachers to take on and share in the responsibility of their own professional development (Sparks & Hirsh, 1997).

The professional development is a key strategy for school success when properly implemented and evaluated. Professional development has and will continue to play an essential role in successful education reform. The question is: what kind of professional development works under a particular context? Based on these evaluation results, we think that one form of high quality professional development can begin with an understanding of scientifically-based reading research; program-specific training during a summer institute; followed by on-going and sustained training supported at the school level with instructional coaches and teacher leaders.

References

Allington, R. L. (2001). *What really matters for struggling readers: Designing research-based programs.* New York: Longman.

Beers, K. (2003). *When kids can't read: What teachers can do*. Portsmouth, NH: Heinemann.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Davey, B. (1983). Think aloud – modeling the cognitive processes of reading comprehension. *Journal of Reading, 27*(1), 44-47.

Desimone, L.M., Smith, T. M., & Ueno, K. (2006). Are teachers who need sustained, content-focused professional development getting it? An administrator's dilemma. *Educational Administration Quarterly, 42* (2), 179-215.

Education Sciences Reform Act of 2002, Pub. L. No. 107-279.

Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children.* Portsmouth, NH: Heinemann.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology, 48*, 163-189.

Graves, M. F. (1999). *Fostering high levels of reading and learning in secondary students.* http://www.readingonline.org/articles/graves1/main.html ERIC Document Reproduction Service (ED 443 090)

Graves, M. F., & Graves, B.B. (1994). *Scaffolding reading experiences: Designs for student success.* Norwood, MA: Christopher-Gordon.

Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.

Guskey, T. R. (2001a). The backward approach. *Journal of Staff Development*, *22*(3), 60.

Guskey, T. R. (2001b). Backward planning: An outcomes-based strategy for professional development. *Curriculum in Context, 28*(2), 18-20.

Hahn, M.L. (2002). *Reconsidering the read-aloud*. Portland, ME: Stenhouse.

Labbo, L. D., & Teale, W. H. (1990). Cross-age reading: A strategy for helping poor readers. *The Reading Teacher,* 362-369.

Muñoz, M. A. (2005). Backward mapping. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (p. 29). Thousand Oaks, CA: Sage.

Muñoz, M. A. (2007). Improving reading in high schools: Outcomes of ramp up to advanced literacy in a large urban district. *Planning and Changing, 38 (1&2),* 89–107.

Pearson, P. D. (1994). Integrated language arts: Sources of controversy and seeds of consensus. In L. Mandel Morrow, J.K. Smith, & L.C. Wilkinson (Eds.), *Integrated Language Arts: Controversy to Consensus* (pp. 11-31). Boston: Allyn and Bacon.

Peterson, C. L., Caverly, D. C., Nicholson, S. A., O'Neal, S., & Cusenbary, S. (2000). *Building reading proficiency at the secondary level: A guide to resources.* Austin, TX: Southwest Educational Development Laboratory.

Richardson, J. (2000). *Read it aloud: Using literature in the secondary content classroom.* Newark, DE: International Reading Association.

Robb, L. (2000). *Redefining staff development: A collaborative model for teachers and administrators.* Portsmouth: Heinemann.

Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A systematic approach* (6th ed.). Newbury Park, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston, MA: Houghton Mifflin Co.

Shellard, E. (2001). Critical issues in developing a high-quality reading program. *ERS Spectrum*, 1-12. http://www.ers.org/spectrum/fall01a.htm.

Sparks, D., & Hirsch, S. (1997). *A new vision for staff development.* Oxford: National Staff Development Council.

U.S. Congress (2001). *No Child Left Behind Act of 2001*. Washington, DC: Author.