IMPROVING READING IN HIGH SCHOOLS: OUTCOMES OF RAMP UP TO ADVANCED LITERACY IN A LARGE URBAN DISTRICT

As emphasized in the recent No Child Left Behind (NCLB, 2001) legislation and the Education Sciences Reform Act (ESRA, 2002), it is critically important to identify "what works" in bringing all students to proficiency levels in core subjects like reading. One philosophy is that enacting effective educational change must involve an entire school rather than a collection of isolated programs. Use of whole-school reform models was boosted substantially by the congressional enactment in 1997 of the Comprehensive School Reform Demonstration (CSRD) program. To date, there are close to 400 models that have been adopted under CSRD support (Desimone, 2002).

Although the philosophy and research-based grounding of CSR approaches should be conducive to schools improving student achievement, the evidence in support of such impacts remains limited (e.g., Berends, Chun, Schuyler, Stockley, & Biggs, 2002; Borman, Hewes, Overman, & Brown, 2003; Herman, 1999). Recently, Borman et al. (2003) conducted a meta-analysis on CSR effects and found an overall achievement advantage for the CSR models examined. Models classified as having the strongest evidence of yielding benefits were Direct Instruction, Success for All, and the School Development Program. Effect sizes were in the +0.15 to +0.21 range, suggesting these models have a modest observed effect. Discouragingly, 17 out of the 29 models were classified as having insufficient statistically reliable or generalizable results to judge effects.

The lack of convincing empirical evidence for an educational program may be more a reflection of limited opportunities available for research on that program than of its ineffectiveness. However, in concert with an emphasis on accountability for schools, NCLB (2001) and ESRA (2002) explicitly advocate the use of rigorous scientifically-based research for determining which educational programs work to raise student achievement (Eisenhart & Towne, 2003; Feuer, Towne, & Shavelson, 2002).

The purpose of the present study was to examine the effectiveness of Ramp Up to Advanced Literacy, an unbundled CSR model, on the reading achievement of ninth grade students in a large urban school district in Kentucky. Using a pre- and posttest impact evaluation design, data from participating and non-participating schools and students were compared. In the current study, only the literacy instruction component of the program was implemented, which is atypical for this program; however, implementing whole-school reform programs is expensive. Given that the literacy component is at the core of a typical comprehensive school program, it remains relevant to describe the Ramp Up program as a CSR model. In the next section, an overview of the current literature on secondary school reading is presented.

Review of the Literature

Research on Reading at the Secondary Level

A voluminous body of research in reading and the broader topic of literacy has been generated by researchers. 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, natural sciences, 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. 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 some high school students, there have been fewer efforts to deal with the problem than to improve reading instruction in the beginning 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 (1999) identified were "teaching for understanding" (Perkins & Blythe, 1993) 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 for the secondary school reading educator seeking to create an effective program. They described four major themes in the literature on improving secondary reading: (a) motivation, (b) decoding print, (c) comprehension, and (d) transactions with texts.

Motivation to read is especially important at the secondary level because older students are more autonomous than younger students and less

likely to engage in academic tasks simply to please teachers. Building motivation to read is a major challenge facing the secondary teacher. A meta-analysis of the general educational literature revealed that verbal praise and feedback on academic tasks increase intrinsic motivation (Cameron & Pierce, 1994). However, this type of motivation is increased by rewarding not merely the completion of a task, but rather high levels of performance. Thus, the secondary school educators seeking to improve reading will build a classroom environment that encourages the positive effect that makes readers persist (Bauman & Duffy, 1997).

Reading requires mastery of technical skills, like decoding text. With such mastery comes fluency, which in turn leads to the self-reinforcing motivation to continue reading for its own sake. Peterson et al. (2000) stated: "two factors, explicit instruction and teacher responsiveness, seem to characterize most successful instructional programs for building decoding skill" (p. 18). In other words, teaching students the skills of decoding—recognition of the alphabet and phonemic awareness—can be a potentially beneficial approach. Furthermore, expert modeling of fluent reading and repeated readings (Chall, 1996) will help struggling secondary school readers.

Language comprehension is essential to both sustaining reading motivation and understanding school subject-matter. Not surprisingly, the larger a student's vocabulary, the better will be his or her reading (Baker, Simmons, & Kameenui, 1995). However, simply memorizing vocabulary words is not as effective as actively learning—making personal connections and seeing words in multiple sources (Blachowicz & Fisher, 2000).

Comprehension is aided by the 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 alouds" (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 hypothetical dialogue 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 including, for example, aesthetic appreciation of language and literary effects.

According to Peterson et al. (2000), effective secondary school reading programs would likely have several characteristics. These include extensive use of scaffolding—teacher assistance to enable 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 its literacy program. In the next section, the program is described.

Ramp Up to Advanced Literacy

Ramp Up is an instructional program designed to improve the reading of low-achieving high school students. The program is a balanced approach to literacy that requires teachers to use many of the techniques used by successful literacy teachers. One component of the program involves explicitly teaching students to use meta-cognitive strategies: seven "habits of proficient readers." For example, teachers encourage (and also model in the classroom) making connections between what is being read and concerns in the everyday lives of the reader. Such connections are called activating schema, and they provide a powerful strategy for readers to make meaning out of what is being read. Another component of Ramp Up is an extensive use of high interest books that are used for instruction. The books can be used with small groups of students (three to six) who are in Literature Circles that read and discuss the same book.

Ramp Up is a two-year course that seeks to accelerate struggling readers who are two or more years behind grade level in English/Language Arts. Typically, Ramp Up would be offered to struggling readers entering high school (e.g., Grade 9). The course assumes that students are able to decode text and are reading at least at a Grade 3 level. The course focuses on ensuring that students make rapid progress toward becoming fluent readers, develop wide vocabularies, and comprehend on-grade level texts through a variety of instructional approaches: independent reading; read-aloud/think-aloud; whole-group and small-group reading and writing instruction; and collaborative learning situations including book discussion groups, partner reading, cross-age tutoring, and service learning. Ramp Up is a research-based program that tries to capture the latest evidence about the most effective approaches to adolescent literacy. The following components comprise a typical Ramp Up class.

Independent reading. Each Ramp Up class opens with independent reading, a time in which students read a book of their own choosing that is at their ability level. Independent reading is an essential part of improving reading proficiency: It is a time for students to develop and improve fluency, and a time to practice strategies they have previously learned (Allington, 2001; Beers, 2003).

Read-aloud/think-aloud. Read-aloud/think-aloud is an opportunity for students to hear a proficient reader make explicit the thoughts she encounters and grapples with as she reads a text (Hahn, 2002; Richardson, 2000). Students are able to observe an excellent reader employing a variety of comprehension strategies: activating background knowledge, visualizing, questioning, determining importance, making inference, summarizing, and monitoring for meaning.

Work period: Whole group and small group reading instruction. During whole group instruction, students learn a variety of reading and writing lessons. For example, students learn how to select a book, give a book talk, and use sticky notes as a tool for reading with purpose. Students are also placed 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 works with the small group around a text that is appropriate to their level. Students work through the text—asking questions, making predictions—with the teacher as guide so that they can successfully apply what they have learned when reading in other settings (Fountas & Pinnell, 1996).

Work period: Writing instruction. Ramp Up students learn a variety of writing procedures and participate in writing craft and skills lessons. Students learn the stages of the writing process, including planning, drafting, editing, conferencing, and revising. Students 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 to their own writing projects.

Writing instruction and experience offer excellent support for those students who struggle with reading. For example, reading our own writing provides the opportunity to learn to read with a critical perspective. According to Pearson (1994), writing provides excellent support for reading comprehension strategies and activities. In addition, writing assignments can be tailored to incorporate the genre pieces for the portfolios required by the state assessment.

Cross-age tutoring. When older students are paired with elementary students for tutoring in reading, they provide a real-world audience for secondary students to practice literacy strategies and behaviors and develop self-confidence. As tutors, secondary students have a chance to model the strategies they have learned in Ramp Up. The secondary student understanding and application of these strategies is enhanced and reinforced by teaching them to a younger student. Reading children's books (the oral performance) to the younger student is documented to increase the fluency and comprehension skills of the older reader (Labbo & Teale, 1990).

Current Study

Research Questions

Schools participating in the program were compared with non-participating schools on a standardized reading test (the CTBS subtest in reading). However, more than reading outcomes were measured. This study provided an opportunity to measure possible side effects (Rossi, Freeman, & Lipsey, 1999) of the literacy instruction component of Ramp Up on several non-cognitive outcomes, including (a) rate of school attendance and (b) rate of out-of-school suspensions.

The following overarching research questions guided the study: (a) What are the results of the program implementation of Ramp Up? (b)

Does Ramp Up impact students' reading achievement? (c) Does Ramp Up result in improved student attendance? (d) Does Ramp Up result in a decrease in discipline problems? and (e) Does Ramp Up affect individual student achievement as measured on standardized reading test scores?

Research Context

In 1990, the Kentucky General Assembly passed the Kentucky Educational Reform Act (KERA) that mandated a complete restructuring of the public elementary and secondary school system in the areas of finance, governance, curriculum, and assessment (Pankratz & Petrosko, 2000). The school district in Kentucky serving as the research site for this study has more than 93,000 students in Grades K–12 and approximately 150 school sites (21 high schools). 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.

Method

This impact evaluation study employed a pre- and posttest design with a treatment and a comparison group (Rossi et al., 1999). The posttest data were derived from 9th grade scores of 12 Ramp Up high schools and 9 comparison high schools. Pretest data were derived from these students' grade 7 and grade 8 scores. Comparisons were schools not using Ramp Up as their literacy intervention and only providing the regular instruction in English. The data at the school level were analyzed using descriptive statistics. Individual student data were analyzed using inferential statistics.

Although the focus of this study was on student academic and non-academic outcomes, monitoring implementation is a key element to avoid "black box" evaluations (Muñoz, 2005). For this reason, on two separate occasions during the 2003/04 school year, district personnel and external technical advisors observed entire periods in each of the Ramp Up classrooms. The America's Choice Ramp Up Implementation Rubric was used to help judge and report the findings. The rubric was developed to assess the implementation fidelity of the program in the treatment schools. As a result of the implementation check, no school was eliminated from the analyses.

All 21 high schools in the district were potential sources of data since the 12 Ramp Up schools and 9 comparison schools add up to 21. However, some analyses were pursued with selected subsets of treatment and control group students that were as equivalent as possible demographically at the baseline year. For the individual student data analyses, a power analysis was conducted to determine the minimum acceptable sample size needed for this research. The method consisted of power analysis criteria established by Cohen (1988) to determine an acceptable sample size for a study involving an analysis of variance (ANOVA) design. The specifications for the power analysis were (Cohen, 1988): (a) a medium effect size (ES = .25); (b) a planned alpha level (α = .05); and (c) a speci-

fied level of power (1 - β = .80). Based on results of the power analysis, the minimum required sample size was found to be 128. As discussed in the data sources section, the groups differed in their average values for several variables prior to the inception of the program.

Participating Schools

Individual student data from 12 Ramp Up high schools and 9 comparison high schools consisted of N = 3082 students. The 12 schools became involved with Ramp Up as part of their effort to increase student achievement. A multi-step process was followed for inclusion of schools in the program. This included a positive vote from members of the school council, which consisted of the principal, three teachers, and two parents. School councils are mandated in Kentucky schools as a result of KERA. They govern many aspects of the school, including selection of curriculum, participation in school improvement programs, and personnel decisions.

Aggregated data on key variables were calculated for both program and comparison schools at the baseline year (2002/03 school year). As can be seen in Table 1, the urban district that served as the research site educates a high percentage of at-risk students with high poverty levels. Ramp Up schools had a larger percentage of at-risk student populations than the school district as a whole. Comparison schools had higher test scores and attendance rates.

Table 1

School Level Data for 12 Ramp Up High Schools and 9 Comparison High Schools at Baseline Year (2002/03)

Group	Poverty (%)	Mobility (%)	Single house-hold (%)	ECE (%)	Attendance (%)	CTBS reading M
Ramp Up average	43.8	12.0	12.6	59.8	90.7	47.8
Comparison average	34.6	6.9	8.4	54.4	93.4	50.1
District average	39.2	9.5	10.5	57.1	92.1	49.0

Note. Poverty was operationalized as participation in the free/reduced lunch program. ECE = Exceptional Child Education (students with disabilities); CTBS = Comprehensive Test of Basic Skills.

Instrumentation

All data from the 2002/03 and 2003/04 school years were abstracted from computerized files provided by the school district that served as the research site. The primary dependent variable used in this study was the Comprehensive Test of Basic Skills (CTBS) in reading, language arts, and mathematics (Kramer, Conoley, & Murphy, 1992). The CTBS is a standardized achievement test that is group-administered at the end of the

school year. In this study, Grade 9 CTBS normal curve equivalent (NCE) and scale scores were used.

At the school level, the independent variable was membership in a Ramp Up or comparison school. At the individual student level, the independent variable was membership in a Ramp Up group or comparison group. Grade 7 Kentucky Core Content Test (KCCT) scale scores in reading were used as previous achievement test scores and to control for initial differences (i.e., covariate) between treatment and control group students.

Since Ramp Up is a reading curriculum based on research on effective practices, the only cognitive outcome was the reading score on the CTBS. However, in an effort to detect possible positive side effects of the program, non-cognitive indicators of attendance and suspensions were also used as dependent variables. Attending to possible side effects is an important element when conducting school-based research.

Findings

Most results have been calculated as descriptive statistical comparisons on school-level data. Gains from the baseline year to the first year of implementation were calculated. The first four sections below report results of school level analyses pertaining to the first four research questions, respectively. The fifth section presents analyses on individual student level data regarding research question 5. Separate analyses were conducted on the data of students who had been in the two lowest achieving groups in 7th grade.

Program Implementation

On separate occasions during the 2003/04 school year, district personnel and external technical advisors were able to observe entire periods in all of the Ramp Up classrooms in order to document the state of program implementation. The America's Choice Ramp Up Implementation Rubric was used to help judge and report the findings. A brief summary appears in the following paragraphs.

Independent reading. Findings indicated that the independent reading component was being faithfully implemented. Students were reading the "just-right" books in all classrooms visited, and in every one there was a quiet reading environment. Fewer than 10% of students were observed disengaged. Teachers were monitoring the progress of the students. Students were also exercising choice in terms of book selection.

Read aloud/think aloud/talk aloud (RA-TA-TA). This component was also being implemented with fidelity and purpose. Teachers were all reading with enthusiasm and most were doing the appropriate strategy applications. The talk alouds were generally lively and enthusiastic, with most students actively participating; however, there was an observed need to improve the proper use of wait time for a student response.

Work period. The work period was largely conducted in a whole group. Few teachers were observed conducting differentiated activities. All classrooms had students working in small groups on the same assignment. Small-group reading instruction was observed in some classrooms, but small-group reading instruction based on assessed needs was not the norm. Assessment for instructional purposes was a work in progress, especially in terms of the systematic use of running records. Writing occurred daily, and students were observed working in their Reader's and Writer's Notebooks in all classes. Student work was displayed in all classrooms. Course artifacts covered the walls in all classrooms.

Cross-age tutoring. Cross-age tutoring was working extremely well in the Ramp Up classrooms. The connections Ramp Up students were making with the youngsters were transformational. Students did lesson plans and reflections for every visit and were held accountable; these sessions were not considered a free day out of the building.

Overall, based on a random sample of observed classrooms ($n = 8$) in the district under study, the Ramp Up implementation was fully operational. In addition, the principals have been supportive beyond expectation. Principals have been part of the regular professional development activities both in the summer institute and during the regular school year.

Impact on School Level Reading Achievement

Table 2 shows the CTBS reading scores for both the Ramp Up and comparison schools from 2002/03 to 2003/04. The aggregated Ramp Up school data showed an increase in CTBS reading scores from the baseline year to the first year of Ramp Up implementation (1.1 NCE gain); in contrast, the aggregated comparison schools' data showed a loss of -1.8 point.

Table 2

School Level CTBS Reading NCE Scores for Ramp Up and Comparison Schools

Group	2002/03 Mean	2003/04 Mean	Change
Ramp Up average ($n = 12$)	47.8	48.9	1.1
Comparison average ($n = 9$)	50.1	48.3	-1.8

Note. Gains = completed first year of implementation (2003/04) minus baseline year (2002/03). Ramp Up students represent about 22.7% of the total school population. CTBS = Comprehensive Test of Basic Skills; NCE = Normal Curve Equivalent.

Impact on Student Attendance

Table 3 shows the school-level attendance for the Ramp Up and comparison schools. Overall, both Ramp Up and comparison schools made no gains in the first year of Ramp Up implementation since the baseline year (losses of 0.1 and 0.3, respectively). Ramp Up schools had smaller losses in attendance than comparison schools had.

Table 3

School Level Attendance Percentages for Ramp Up and Comparison Schools

Group	2002/03 Attendance rate	2003/04 Attendance rate	Change in attendance rate
Ramp Up average	90.7	90.6	-0.1
Comparison average	93.4	93.1	-0.3

Note. School attendance rates = [(membership in school – absences) / membership in school] x 100. Gains = completed first year of implementation (2003/04) minus baseline year (2002/03).

Impact on Discipline Problems

From the baseline year (2002/03) to the first year of Ramp Up implementation (2003/04), the Ramp Up schools did not decrease the number of out-of-school suspensions. However, the Ramp Up schools had a smaller average increase in out-of-school suspensions compared to comparison schools (27 versus 37, respectively). Table 4 displays discipline data for both Ramp Up and comparison schools.

Table 4

School Level Out-of-School Suspensions Rate per Every 1,000 Students in Ramp Up and Comparison Schools

Group	2002/03 Suspension rate	2003/04 Suspension rate	Change in suspension rate
Ramp Up average ($n = 12$)	283	310	27
Comparison average ($n = 9$)	155	192	37

Impact on Individual Standardized Reading Test Scores

Analyses of 2003/04 CTBS subtests compared Ramp Up and comparison students who had been classified Novices in reading on the Grade 7 state reading assessment (popularly known as the CATS reading test). Novice is the lowest performance category in the state assessment system. The focus was on the CTBS subtests of reading, language arts, and mathematics.

Each of four demographic variables was dichotomized using dummy coding procedures (Pedhazur, 1982). Ethnicity was categorized: (a) minority, and (b) White. The category minority was 92% African American. Free or reduced-price lunch status was categorized as (a) free or reduced-price, or (b) pay for lunch. Single parent homes were categorized as (a) single parent or (b) not single parent. Gender was categorized as (a) male or (b) female.

Chi-square analyses were conducted to determine if the proportion of students in each of the demographic categories (i.e., minority, free

or reduced-price, single parent, female) was equal in the program group (Ramp Up and comparison). The analyses failed to show statistically significant differences between treatment and control groups in: (a) percent minority; (b) percent free or reduced-price lunch status; (c) percent single-parent homes; and, (d) percent females.

Since significant differences were found between the Ramp Up and comparison students on previous achievement, a MANCOVA was performed using previous achievement as the covariate. The independent variable was program group (Ramp Up and comparison) and the dependent variables were three CTBS scale scores in reading, language, and mathematics. Following the suggestion of Stevens (2002), the relationship between the covariate and the dependent variables was tested. A significant multivariate relationship was found, Wilks' $\Lambda = 0.93$, $F(3, 404) = 10.56$, $p < .001$. Thus, there was a justification to use the covariate in the analysis.

The researchers then tested the homogeneity of regression hyperplanes assumption of MANCOVA (Stevens, 2002). The treatment by covariate interaction effect was not significant, Wilks' $\Lambda = 0.99$, $F(3, 403) = 0.79$, $p = .79$. Thus, the MANCOVA could proceed without a bias in the test of treatment effects due to non-parallel regression hyperplanes.

Descriptive unadjusted and adjusted statistics for each group on the three CTBS subtests are summarized in Table 5. As can be seen, the Ramp Up mean was directionally higher than the comparison group mean on reading and language arts; however, given that Ramp Up is a literacy program, an impact on mathematics was not found. The MANCOVA was statistically significant, Wilks' $\Lambda = 0.96$, $F(3, 404) = 5.33$, $p < .01$. Thus, there were differences among the Ramp Up and non-Ramp Up subjects on adjusted means for the three CTBS scores that were used as dependent variables. Follow-up univariate tests were significant on reading [$F(1, 404) = 10.22$, $p = .001$] and language [$F(1, 404) = 6.40$, $p = .01$], but not in mathematics [$F(1, 404) = .17$, $p = .69$]. As indicated in Table 5, the effect sizes associated were small to large in magnitude ranging from -0.04 to +0.30.

Table 5

Unadjusted Means and Standard Deviations for Novices in Ramp Up and Matched-Control Students on CTBS Ninth-Grade Scale Scores in 2003/04 (Year 1)

Comparison group and subtest	M	M_{adj}	SD	Effect size[a]	Effect size[b]
Language					
Ramp Up	631.81	631.21	36.67	+0.28	+0.25
Control	621.22	621.80	38.97		
Reading					
Ramp Up	641.16	640.41	27.86	+0.35	+0.30
Control	630.74	631.47	30.26		

(continued)

Table 5 *(continued)*

Comparison group and subtest	M	M_{adj}	SD	Effect size[a]	Effect size[b]
Mathematics					
Ramp Up	638.38	637.98	42.69	-0.02	-0.04
Control	639.25	639.64	40.38		

Note: Ramp Up *n* = 201; Control *n* = 208. CTBS = Comprehensive Test of Basic Skills. Effect sizes[a] were computed from the unadjusted means. Effect sizes[b] were computed from the adjusted means.

    Given that sample size was enough to conduct another individual student analysis, a second study was designed to assess the impact of Ramp Up on the second lowest level of academic performance on the Grade 7 state reading assessment (i.e., Low Apprentice). To determine group equivalence on the possible covariates, we first conducted chi-square analyses comparing the Ramp Up and comparison groups; no statistically significant differences were found in percent minority, females, free/reduced-price lunch, single-parent homes, or in previous achievement. Results of a MANOVA indicated no significant differences between Ramp Up and non-Ramp Up Apprentice students in reading, language arts, and mathematics. As can be seen in Table 6, the Ramp Up Apprentice students' averages were directionally lower than the comparison group mean on reading, language arts, and mathematics. The MANOVA Wilks' Λ yielded a non-significant overall comparison group effect, $F(3, 295) = 1.19$, $p = .31$. Follow-up univariate tests were not significant on reading [$F(1, 295) = .16$, $p = .69$], language arts [$F(1, 295) = 1.12$, $p = .29$], and mathematics [$F(1, 295) = 3.00$, $p = .08$]. Descriptive statistics and effect sizes for each group on the three CTBS subtests are summarized in Table 6.

Table 6

Means and Standard Deviations for Apprentice Ramp Up and Matched-Control Students on CTBS Ninth-Grade Scale Scores in 2003/04 (Year 1)

Comparison group and subtest	M	SD	Effect size
Language			
Ramp Up	644.80	31.40	-0.12
Control	649.04	37.12	
Reading			
Ramp Up	656.24	25.28	-0.05
Control	657.56	30.76	
Mathematics			
Ramp Up	653.47	44.49	-0.20
Control	662.45	44.83	

Note: Ramp Up *n* = 139; Control *n* = 160. CTBS = Comprehensive Test of Basic Skills.

A third and final analysis of 2003/04 CTBS subtests compared Ramp Up and comparison students who had been Novices and Low Apprentices when they were measured on the Grade 7 state reading assessment. These were the two lowest performance categories in the state assessment system. Again, the focus was on the CTBS subtests of reading, language arts, and mathematics. Initial analyses, conducted to verify the similarity of the program groups, failed to show significant differences on previous achievement scores and some of key demographic variables examined, including free or reduced-price lunch, single parent homes, and females. Since the Ramp Up group had more minority students, a MANCOVA was employed to use percent minority as a covariate. Following the suggestion of Stevens (2002), the relationship between the covariate and the dependent variables was tested. A significant multivariate relationship was found, Wilks $\Lambda = 0.98$, $F(3, 703) = 4.85$, $p < .001$. Thus, there was a justification to use the covariate in the analysis.

Descriptive statistics for each group on the three CTBS subtests are summarized in Table 7. As can be seen, the Ramp Up mean was directionally higher than the comparison group mean on reading, but not in language arts and mathematics. The MANCOVA was statistically significant, Wilks $\Lambda = 0.98$, $F(3, 703) = 4.13$, $p < .01$. Follow-up univariate tests were significant only in reading [$F(1, 703) = 5.64$, $p = .02$], but not in language arts [$F(1, 703) = 2.34$, $p = .13$] and mathematics [$F(1, 703) = 1.38$, $p = .24$]. As indicated in Table 7, the effect sizes associated with reading and language arts ranged from -0.09 to +0.18.

Table 7

Unadjusted and Adjusted Means and Standard Deviations for Novices and Low Apprentices in Ramp Up and Matched-Control Students on CTBS Ninth-Grade Scale Scores in 2003/04 (Year 1)

Comparison group and subtest	*M*	M_{adj}	*SD*	Effect size[a]	Effect size[b]
Language					
Ramp Up	637.12	637.42	35.15	+0.10	+0.12
Control	633.31	633.04	40.55		
Reading					
Ramp Up	647.33	647.62	27.81	+0.16	+0.18
Control	642.40	642.14	33.22		
Mathematics					
Ramp Up	644.55	645.03	44.00	-0.11	-0.09
Control	649.34	648.89	43.85		

Note: Ramp Up *n* = 340; Control *n* = 368. CTBS = Comprehensive Test of Basic Skills. Effect sizes[a] were computed from the unadjusted means. Effect sizes[b] were computed from the adjusted means.

As a further exploration of the data from Novices and Low Apprentices, the researchers tested the homogeneity of regression hyperplanes assumption of MANCOVA. The treatment by covariate interaction effect was significant, Wilks' $\Lambda = 0.99$, $F(3, 703) = 2.87$, $p = .04$. The partial eta square statistic for this effect was .012, meaning the effect size was small.

To further explore the data, a two-way MANOVA was performed with the independent variables ethnicity (minority, White) and treatment group (Ramp Up, comparison). Dependent variables were the three CTBS scale scores. All three effects of the MANOVA were significant: (a) for the main effect of ethnicity, Wilks' $\Lambda = 0.98$, $F(3, 702) = 4.37$, $p = .01$; (b) for the main effect of treatment conditions, Wilks' $\Lambda = 0.98$, $F(3, 702) = 4.11$, $p = .01$; and (c) for the interaction effect of ethnicity by treatment conditions, Wilks' $\Lambda = 0.99$, $F(3, 702) = 2.87$, $p = .04$.

The individual dependent variables were examined. Regarding the main effect of ethnicity, there were significant differences favoring White versus minority students for all three CTBS tests (reading, language arts, mathematics, all comparisons at $p < .02$). For the main effect of instructional condition, the effect of Ramp Up on CTBS reading was not significant at $p = .06$; however the effects on language arts ($p = .18$) and mathematics ($p = .09$) were farther away from the $p = .05$ criterion of significance.

An interesting finding was the ordinal interaction effects found for reading. The interactions show less of an achievement gap between ethnic minority and White students in the Ramp Up condition than in the comparison condition. On CTBS reading, a significant interaction occurred, $F(1, 704) = 4.40$, $p = .04$. For comparison subjects, the mean for Whites, $M = 648.81$, exceeded the mean for minority students, $M = 637.42$, a gap of 11.4 points. However, for Ramp Up subjects, the mean for Whites, $M = 648.29$, only slightly exceeded the mean for minority students, $M = 646.79$, a negligible gap of 1.5 points. Figure 1 displays the group (Ramp Up, comparison) by race (minority, White) interaction plot in CTBS reading.
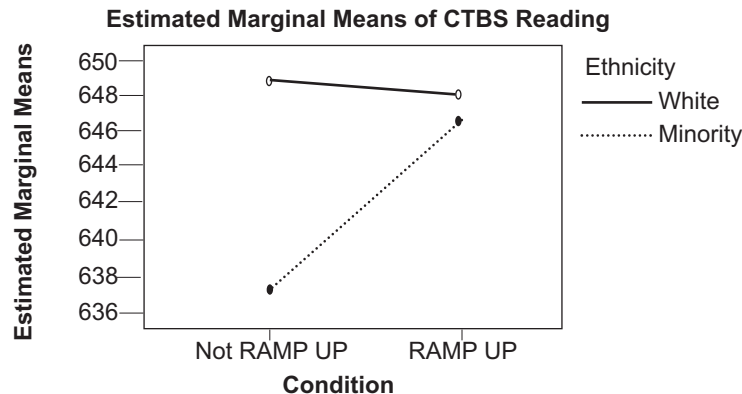

Figure 1. Interaction of ethnicity and condition on CTBS reading for Novice and Low Apprentice students (Ramp Up, $n = 340$; Not Ramp Up, $n = 368$). CTBS = Comprehensive Test of Basic Skills.

Discussion

In an effort to assist schools in making instructional and organizational changes, an abundance of schoolwide reform models have emerged. The concept of Ramp Up is to restructure 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. Data analysis showed that the model had a positive effect on CTBS reading and language arts score gains. The results are generally supportive of positive Ramp Up effects. On two of the three CTBS subtests in ninth grade, Ramp Up Novice students significantly surpassed the matched comparison students. As expected, however, no differences were found on the mathematics subtest of the CTBS. Furthermore, the effect sizes for language arts and reading (range = +0.25 to +0.30) for the statistically significant comparisons were not modest in size and surpassed the levels associated with the more successful CSR models identified in the research literature (Borman et al., 2003). No significant positive effect was observed, however, on the next group of low academic performance students (Low Apprentice). The combined sample showed a positive effect of Ramp Up in reading, but not in language arts or math. An interaction effect indicated that Ramp Up had a positive effect in closing the achievement gap between minority and White students in reading.

Based on site visits and conversations with staff associated with the Ramp Up schools, there is evidence that any positive effects of the program on Novices are partly a consequence of the quality of implementation, strong external technical assistance, and district-level support. The good quality of implementation in the district that served as research site may be associated with the large degree of teacher buy-in and involvement with the comprehensive school reform change (Sarason, 1990, 1996). The service providers have made available technical assistance that involves (a) regular visits to each participating school and (b) annual reports of implementation fidelity. The district provides resource teachers and multiple opportunities for professional development.

Given the narrow curriculum focus and intensity of high stakes testing environments (Linn, 2003; Shepard, 2000), a concern might be whether student achievement on state-mandated basic-skills tests could be noticeably impacted after only one year (see Desimone, 2002; Fullan, 2000; Levin, 1993). Given the prescriptive nature of Ramp Up, less time is needed for successful implementation. This is the opposite of process-based models (Datnow, Hubbard, & Mehan, 2002). Given the above rationale, it is noteworthy that after only one year of implementing Ramp Up, the lowest performing students demonstrated a significant overall advantage (effect size = +0.28 on average) over comparison students on ninth-grade CTBS Reading and Language Arts.

In the present study, a comprehensive approach to effect size interpretation was used (Cooper, 1981). Most generally, the researchers compared the overall Ramp Up effect size to Cohen's (1988) definitions

of a small effect ($d = 0.20$) and a large effect ($d = 0.80$). Cohen (1988) pointed out that the relatively small effects of around $d = 0.20$ were most representative of fields that are closely aligned with education. Similarly, and more specifically, Lipsey and Wilson (1993) organized a compendium of meta-analyses in which they concluded that educational treatment effects of modest values, of even $d = 0.10$ to $d = 0.20$, should not be interpreted as trivial.

Even more specifically, Ramp Up effects were compared to previous national efforts to help improve the outcomes of large numbers of high-poverty and low-achieving students and schools. One of the most obvious comparisons to the effect of CSR is the effect of traditional Title I programs. Title I programs were the subject of a meta-analysis (Borman & D'Agostino, 1996) that synthesized the results from all federal evaluations conducted between 1965 and 1994. Borman and D'Agostino (1996) estimated that the average effect size associated with these efforts was $d = 0.11$. In the present study, the effect sizes were superior to those in Title I programs and were at the same level of the more successful CSR models identified in the research literature (Borman et al., 2003).

Why and how student learning improved are less clear based on the present data. Assuming the implementation fidelity to be valid, the combination of improved student engagement and enthusiasm for learning coupled with high teacher involvement would certainly be predictive of observable achievement gains. However, as shown in prior research with New American School models, specifically in Memphis (Ross, 2001) and San Antonio (Berends et al., 2002), early success may be realized as part of extraordinary commitment to and expectations for reform.

The key factor, which seems to have plagued most prior systemic reform efforts, is demonstrating sustainability over time. Subsequent research is therefore needed to follow up these early, promising results with the unbundled Ramp Up model. To extend the present study and answer some important questions that still remain, such studies should involve rigorous examination of model implementation and process outcomes via triangulated data sources reflecting participant/stakeholder perceptions, activities, and classroom practices. It would be valuable to study differences in achievement between minority students and White students. Any program that reduces the achievement gap between different ethnic groups, as this study suggests Ramp Up did, should be further studied.

Future researchers need to study how teachers across the nation are helping struggling high school readers. Literacy programs like Ramp Up are helpful to meet this need of training for secondary teachers. The root of the problem, however, is that many secondary teachers never anticipated needing to help students learn to read. As we know, many teachers with secondary certificates had very limited coursework in the content area of reading in their teacher preparation programs at colleges and universities. Colleges and universities preparing future educators must redouble their efforts to address the need associated with struggling readers in our public secondary schools (Beers, 2003).

References

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

Baker, S. K., Simmons, D. C, & Kameenui, E. J. (1995). *Vocabulary acquisition: Synthesis of the research.* (Technical Report No. 13). Eugene, OR: University of Oregon, National Center to Improve the Tools of Educators.

Bauman, J. F., & Duffy, A. M. (1997). *Engaged reading for pleasure and learning: A report from the National Reading Research Center.* Athens, GA: National Reading Research Center.

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

Berends, M., Chun, J., Schuyler, G., Stockley, S., & Briggs, R. J. (2002). *Challenges of conflicting school reforms: Effects of New American Schools in a high-poverty district.* Santa Monica, CA: Rand.

Blachowicz, C. L. Z., & Fisher, P. (2000). Vocabulary instruction. In M. L. Kamil, P. B. Rosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of Reading Research*, Vol. III. Mahwah, NJ: Lawrence Erlbaum.

Borman, G. D., & D'Agostino, J. V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis, 18,* 309–326.

Borman, G. D., Hewes, G., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125–230.

Cameron, J., & Pierce, J. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research, 26,* 363–423.

Chall, J. (1996). *Stages of reading development.* New York: McGraw-Hill.

Cohen, J. (1988). *Statistical power analysis for the social sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Cooper, H. (1981). On the effects of significance and the significance of effects. *Journal of Personality and Social Psychology, 41,* 1013–1018.

Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending educational reform: From one school to many.* New York: Routledge-Falmer.

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

Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72,* 433–480.

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

Eisenhart, M., & Towne, L. (2003). Contestation and change in national policy on "scientifically based" education research. *Educational Researcher, 32*(7), 31–38.

Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher, 31*(8), 4–14.

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

Fullan, M. (2000). The return of large scale reform. *Journal of Educational Change, 1,* 5–28.

Graves, M. F. (1999). *Fostering high levels of reading and learning in secondary students.* (ERIC Document Reproduction Service No. ED443090). Retrieved April 30, 2007, from http://www.readingonline.org/articles/graves1/index.html

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

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

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

Herman, R. (1999). *An educators' guide to schoolwide reform.* Arlington, VA: Educational Research Service.

Kramer, J. J., Conoley, J. C., & Murphy, L. L. (1992). *The eleventh mental measurements yearbook.* Lincoln, NE: University of Nebraska.

Labbo, L. D., & Teale, W. H. (1990). Cross-age reading: A strategy for helping poor readers. *The Reading Teacher, 43*(6), 362–369.

Levin, H. (1993). *Learning from accelerated schools.* Unpublished paper, Stanford University, Palo Alto, CA.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectancies. *Educational Researcher, 32*(7), 3–13.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48,* 1181–1209.

Muñoz, M. A. (2005). Black box. In S. Mathison (Ed.), *Encyclopedia of Evaluation* (pp. 34–35). Thousand Oaks, CA: Sage.

No Child Left Behind Act of 2001, Pub. L. No. 107-110.

Pankratz, R. S., & Petrosko, J. M. (2000). *All children can learn: Lesson from the Kentucky reform experience.* San Francisco: Jossey-Bass.

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

Pedhazur, E. J. (1982). *Multiple regression in behavioral research: Explanation and prediction.* New York: CBS College.

Perkins, D., & Blythe, T. (1993). Putting understanding up front. *Educational Leadership, 51*(5), 4–7.

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. (ERIC Document Reproduction Service No. ED458562)

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

Ross, S. M. (2001). *Creating critical mass for restructuring: What we can learn from Memphis.* Charleston, WV: AEL Policy Briefs.

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

Sarason, S. (1990). *The predictable failure of educational reform.* San Francisco: Jossey-Bass.

Sarason, S. (1996). *Revisiting the culture of the school and the problem of change.* New York: Teachers College Press.

Shellard, E. (2001). Critical issues in developing a high-quality reading program. *ERS Spectrum, 19*(4), 4–11. Retrieved April 23, 2007, from http://www.ers.org/spectrum/fall01a.htm

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.

Marco A. Muñoz is an Evaluation Specialist in the Accountability, Research, and Planning Department of the Jefferson County Public Schools, Louisville, Kentucky.