RICHARD M. KUBINA JR., The Pennsylvania State University, MICHAEL LAMPORT COMMONS, Harvard Medical School, and BARBARA HECKARD, Lebanon School District

Using Precision Teaching with Direct Instruction in a Summer School Program

Abstract: This study examined the effects of a Direct Instruction (DI) reading program combined with Precision Teaching during a public school's summer program. Students received instruction from *Reading Mastery* programs for a six-week period. Students also practiced specific reading skills including letter-sound identification, sounding out words, and passage decoding, and they displayed their data on Standard Celeration Charts. Results showed that DI combined with Precision Teaching produced statistically significant gains as measured by informal and formal tests of reading. The results also indicated small to moderate effect sizes for the reading measures.

The climate for research-based, or evidence-based, approaches for reading has changed. The No Child Left Behind Act of 2001 (NCLB) mandated evidence-based criteria as a standard for judging which programs will receive federal education funding. Evidence-based programs have undergone scientific testing and have yielded reliable and valid results. By incorporating evidence-based programs, the educational outcomes of students across the nation will improve (Whitehurst, 2002). Indeed, without the use of research-based practices as a guide, true reform efforts in education are unlikely to occur (National Research Council, 1998).

In recent years, a number of evidence-based programs have been developed and tested to put the educational reform effort into action. For example, DI reading programs, developed by Engelmann and colleagues, have consistently and reliably shown significant success with diverse groups of learners (Adams & Engelmann, 1996; Carnine, Silbert, Kame'enui, & Tarver, 2004; Hempenstall, 2004; Kinder, Kubina, & Marchand-Martella, 2005; Marchand-Martella, Slocum, & Martella, 2004). The range of learners extends from students in special education to those in gifted programs (Adams & Engelmann). These developmental reading programs include *Horizons*, *Journeys*, *Reading Mastery Classic*, and *Reading Mastery Plus*, whereas remedial programs consist of *Corrective Reading—Decoding* and *Comprehension*. DI reading programs form a comprehensive curriculum with teacher presentation manuals, student books, and other materials.

DI also has some activities built into the programs to gauge student progress. Some of these progress mechanisms include skills-profile folders and mastery test checkouts. However, DI programs may further benefit from an additional standard graphic display system and a standard set of graphing conventions for student progress. One such classroom-based procedure that helps to

Journal of Direct Instruction, Vol. 9, No. 1, pp. 1-12. Address correspondence to Richard Kubina at rmk11@psu.edu.

measure student behavior with standard charts and conventions and facilitates curricular decisions, Precision Teaching, may further augment the educational outcomes produced by DI.

Lindsley (1997) defined Precision Teaching as a set of tactics and strategies that assist with the analysis and interpretation of behavior. Precision Teaching uses a Standard Celeration Chart to display data in a formative manner. Teachers have used Precision Teaching in both public- and private-school classrooms as well as with a variety of learners spanning various ages, genders, and disabilities (Johnson & Layng, 1992; Kubina & Morrison, 2000; Lindsley, 1990, 1997; Mercer, Mercer, & Evans, 1982; West & Young, 1992). Precision Teaching, like DI, meets the criteria for an evidence-based approach to education.

However, unlike DI, Precision Teaching does not specify what or how to teach. It offers a method to measure behavior, display the data on the Standard Celeration Charts, and facilitate decision-making for a teacher. Precision Teaching has four important guidelines that influence its use: (a) a focus on directly observable behavior, (b) the use of frequency as a standard unit of measurement, (c) data displayed on a Standard Celeration Chart, and (d) the belief that the "learner knows best" or the practice of embracing data as a reflection of the current environmental effects influencing a behavior (Kubina, Ward, & Mozzoni, 2000; White, 1986, 2005).

As shown by previous research, teachers who formatively assess students and use graphs make more responsive decisions than teachers who do not (Fuchs & Fuchs, 1986). The specialized Standard Celeration Chart and systematic practice procedures used in Precision Teaching have facilitated improved outcomes in both public and private schools (Beck & Clement, 1991; Johnson & Layng, 1992, 1994; Maloney, 1998). As a public school example, Sacajawea Elementary in Great Falls, MT, implemented Precision Teaching throughout the school. The intervention consisted of teachers adding approximately 20 to 30 minutes of daily practice, Standard Celeration charting, and subsequent decision-making. Aggregated achievement test scores increased an average of 20 to 40 percentile points from the previous level after 3 years of the Precision Teaching intervention (Beck & Clement, 1991; Binder & Watkins, 1989). By combining Precision Teaching and DI, teachers and students may experience additional benefits. As Binder and Watkins (1990) put it, "Precision Teaching and Direct Instruction are mature and extremely powerful instructional technologies that are fully capable of erasing America's 'basic skills crisis' if widely adopted" (p. 93).

Further research exists supporting the proposition that DI reading programs show positive results when combined with Precision Teaching techniques (Blackwell, Stookey, & McLaughlin, 1996; Edmonson, Peck, & McLaughlin, 1996; Haring & Krug, 1975; Holz, Peck, McLaughlin, & Stookey, 1996; Johnson & Layng, 1992; Johnson & Street, 2004; Maloney, 1998; Morrell, Morrell, & Kubina, 1995; Neely, 1995; Stenseth & McLaughlin, 1996). For instance, Morrell et al. examined the effects of practicing sight words from *Reading Mastery I* with three second-grade students who had specific learning disabilities in reading. An instructional day consisted of following the *Reading Mastery I* curriculum as well as supplementing 5 to 10 minutes of systematic practice and Standard Celeration charting of the data. The intervention helped students to proceed through the lessons rapidly and improved their reading of targeted words within sentences. The students began the intervention of *Reading Mastery* and Precision Teaching toward the end of the school year and could fluently read more than 40 sight words from *Reading Mastery I* within 2 months. Prior to the DI and Precision Teaching intervention, the students could not read any words.

By adding Precision Teaching to DI programs, teachers have a powerful combination of carefully designed instruction and a "sophisticated set of measurement practices" and "productive practice exercises" capable of producing substantial academic gains (Desjardins & Slocum, 1993, p. 20). Considering the critical need for producing competent readers, combining DI reading programs with Precision Teaching may have a positive synergistic effect. To date, published articles describing large-scale combinations of Precision Teaching and DI in a public school do not exist. Therefore, in this study we examined the effects of a public school district's summer school program that combined the DI reading program *Reading Mastery* with Precision Teaching.

Method

Participants and Setting

The participants came from an urban district located in central Pennsylvania. There were 203 students, including 89 girls (43.8%) and 114 boys (56.2%), from five elementary schools attending the summer school program. Selection criteria for summer school included scoring at the 25th percentile or lower on the *Pennsylvania System of School Assessment* (Pennsylvania Department of Education) and performing below grade level in reading. There were 61 first graders (30.0%), 53 second graders (26.1%), 49 third graders (24.1%), and 40 fourth graders (19.7 %).

Of the student population for summer school, 36 (18%) were identified, using standardized methods, as "Limited English Proficient, LEP," and 26 (13%) of the students had an Individualized Education Program. All of the eligible students in each participating classroom participated in the study. The students attended summer school, which ran 4 days a week for 6 weeks. Class size ranged from 10 to 14 students per class. Each class had a teacher and a paraprofessional.

Materials

Reading Mastery Rainbow Editions I, II, and *III* were used (Engelmann & Bruner, 1995a; Engelmann & Bruner, 1995b; Engelmann & Hanner, 1995). Each *Reading Mastery* program contained a teacher presentation book, student reading books, and student workbooks. To display daily reading practice measures, Standard Celeration Charts were used. Other materials included practice sheets, pieces of Mylar, dry-erase pens, and paper towels. All students used a data sheet to record practice scores before displaying them on Standard Celeration Charts.

Response Measurement

The difference between each student's pretest and posttest measure served as the method to evaluate the results of *Reading Mastery*, the selected skills practiced to fluency, and the Standard Celeration charting methods from Precision Teaching. During the first week of summer school, before students received instruction, a team of principals, school psychologists, and teachers administered the pretest measures. During the last week of summer school, the same team administered the posttest measures to the students. The assessors gave three informal and three formal measures of reading.

Informal measures of reading. Classroom teachers and paraprofessionals implemented the informal measures (these assessments are available from the first author upon request). For all three informal measures, assessors gave directions, modeled the performance, and asked if the student understood. When students did not understand a direction, the assessors repeated the direction, modeled the performance, and led the students to the correct response. During each informal reading measure, the assessors scored correct and incorrect answers out of the students' view. If students made mistakes, they did not receive feedback on their errors. Additionally, if students hesitated for more than five seconds on any part of

the informal measures, the assessors told the students the correct response, marked it as incorrect, and told the students to keep going. At the end of each informal measure, the assessors made positive comments and thanked the students for participating.

Letter sound fluency measure. The first informal measure of reading required students to point to and say as many letter sounds as they could in 1 minute. The letter-sound sequence came from the *Reading Mastery I* teachers' guide (Engelmann & Bruner, 1995c). A sheet set in landscape view had the first 40 letter sounds in random order and then repeated the order two more times for a total of 120 letters. Macrons were used to show the long sounds. The assessor gave the student instructions and then modeled how to point to a letter, say its sound, and move across the page in a left-to-right fashion. After asking if the student had any questions, the assessor told the student to begin and started the timer, which was set for 1 minute. At the end of the timing, the student left and the assessor scored and recorded the student's performance.

Orally decoding words fluency measure. The second informal measure required students to sound out words and then say them fast. All words came from *Reading Mastery I, II,* and *III* sight-word lists and were taken from advanced parts of each program. It was possible that some students in advanced *Reading Mastery* lessons (e.g., *RM III*) had already been taught some of the words (e.g., *RM I*). The regular words had a mixture of word types (e.g., C = Consonant; V =Vowel: CV, VC, CVC, CVCC, CCVCC) and words beginning with continuous and quick sounds. Each sheet had a total of 60 words. The assessors provided directions and modeled how to sound out words and then say them fast. To record correct and incorrect answers, a separate sheet was used to follow along with the students. The assessors awarded the students one point for each correctly identified letter sound and one point for saying the word fast. For instance, the word "run" had a potential score of four with one correct point for each letter sound and one point for saying the word fast.

Oral reading fluency measure. The third informal test measured how many correct words per minute the students read. Assessors used a story taken from a lesson at the end of the storybook, depending on which *Reading Mastery* program each student tested into. For example, if a student placed into *Reading Mastery II*, she read a passage from Lesson 60 during both the pretest and posttest. The passage was selected from a lesson that the students would not read before summer school ended. For each of the informal measures, students could have encountered sounds and words not yet instructed.

Formal measures of reading. Three subtests from the *Woodcock Reading Mastery Test-Revised-NU* (Woodcock, 1998) served as formal measures of reading. The subtests, "Word Attack" and "Word Identification," provided formal measures of the students' skills in correctly pronouncing words and employing analytic decoding strategies. The other subtest, "Passage Comprehension," gave information regarding the students' skills in comprehending what they read. Only three assessors (i.e., one principal and two school psychologists), who were trained to administer the subtests from the *Woodcock Reading Mastery Test-Revised-NU*, assessed the students. The assessors administered Form G for the pretest and Form H, which had parallel test items, for the posttest.

Research Design

To examine the effects of the combination of *Reading Mastery* and Precision Teaching, the investigators used a pre-experimental, one-group pretest-posttest design (Fraenkel & Wallen, 1996). The one-group pretest-posttest design, however, contains a number of threats to internal validity. As Fraenkel and Wallen point out, any of the nine identified threats to internal validity could explain the results of the posttest. Therefore we recom-

mend that readers interpret the subsequent results cautiously.

Procedure

Each classroom had a teacher who taught *Reading Mastery* lessons to homogeneously grouped students. All teachers had previously taught *Reading Mastery* for a minimum of 1 year. The teachers also attended a district training aimed at providing additional instruction for the summer school. At the district training, both teachers and paraprofessionals learned to implement certain aspects of Precision Teaching for the summer school program. Specifically, they learned how to use the Standard Celeration Chart and how to set up practice-to-fluency activities for letter sounds, oral decoding of words, and passage reading (Kubina, 2005). The initial training, conducted by the first author, spanned two days and occurred prior to summer school. Throughout the six weeks of summer school, the teachers received periodic coaching sessions. Coaching sessions involved checking data on the Standard Celeration Chart, reviewing instructional decisions, and answering any teacher questions. During summer school, teachers who used *Reading Mastery* continued to use the program as they were trained and did not change any formats or instructional delivery techniques.

Letter sounds. A sheet with letter sounds was used to practice saying letter sounds fluently. Five sheets of letter sounds were used depending on the students' current level of instruction in *Reading Mastery*. Sheet A consisted of the first 8 letter sounds from the *Reading Mastery* letter-sound sequence placed in random order on 8 x 11 in. landscape-view paper. The letters filled the page and appeared in equal proportion. Therefore, if a letter sound sheet had 120 total sounds, each separate letter sound appeared 15 times. Sheets B, C, D, and E each added another 8 sounds, so that B had 16 letter sounds from the *Reading Mastery* sequence, and C, D, and E contained 24, 32, and 40 letter sounds respectively. All letter sounds followed the previously mentioned instructional design of using an 8 x 11 in. landscape view of 120 letters per page.

Each day, students practiced saying their letter sounds with a partner who was also a student. The students engaged in practice as a group. The teacher started a countdown timer and told the students when to begin and when to stop. Students were taught how to record correct and incorrect answers on a sheet and then to provide feedback to the partner. After providing feedback to the partner, students switched roles so that all students had an opportunity to practice each day. The Precision Teaching fluency aim for letter sounds was 100 to 120 letter sounds per minute (Freeman & Haughton, 1993). First-grade students practiced for 20 seconds instead of 1 minute and had a goal or fluency aim of 33 to 40 letter sounds per 20 seconds. The goal of 33 to 40 letter sounds was calculated by dividing 60 seconds or 1 minute by three because there are three, 20-second intervals per minute. The second through fourth graders had to reach the fluency aim of 100 to 120 letter sounds per minute. If students struggled with reaching their aim, the teacher could lower the counting time to 30 seconds (i.e., aim would then equal 50 to 60 letter sounds per 30 seconds) or to 20 seconds. Reducing the time interval of practice was an attempt to help the students build endurance, or the ability to perform stably for a given period of time (Binder, 1996). If students were fluent with letter sounds, evidenced by meeting the fluency aim, they did not engage in the practice procedure.

It should be noted that students did not practice letter sounds without first receiving instruction. Because all students were in small groups and received the same instruction, practice did not begin until after the lesson that contained the last letter sound of a sheet. For example, in *Reading Mastery I* the eighth letter sound /i/ was introduced in Lesson 34.

Students practiced sheet A only after passing Lesson 34. Practice continued until a student met the fluency aim. Sheet B was introduced after Lesson 64. Students who mastered letter-sound sheet A before the next letter sheet was introduced were helpers who counted corrects and incorrects or provided help or encouragement directed by the classroom teacher.

Orally decoding words. As described in the second informal measure, students practiced sounding out words and saying them fast. The words came from the word list used in their current *Reading Mastery* program and not from the words used in the informal measure, thus avoiding an overlap. A student on Lesson 20 of *Reading Mastery II* practiced words made up of letter sounds previously instructed. Each sheet had more words than the students could sound out and say fast in a minute. Each *Reading Mastery* program (*I, II,* and *III*) included five different sheets made up of words from 20 lessons, and some words were repeated on the sheet.

Students were taught how to record correct and incorrect answers on the word-list sheet that their partners were using. Partners started from a different place on the word list each time to avoid repeating what the other partner had previously sounded out and then said fast. Because the Precision Teaching published literature did not include fluency aims for orally decoding words, the first author sampled a group of young adults who were considered fluent (Kubina, 2003). The sampling procedure followed the guidelines from Binder (1996) and Koorland, Keel, and Ueberhorst (1990). The fluency aim for second through fourth graders was 80 to 100 letters sounded out and words said quickly per minute. First-grade students used a 20-second counting time with a fluency aim of 27 to 33 letters sounded out and words said correctly. The first-grade students' counting time was calculated by dividing three (i.e., three 20-second intervals in one minute) into the 80 to 100 fluency aim. As an intervention and at the discretion of the teacher, teachers used 20- and 30-second counting times with the second- though fourth-grade students (i.e., fluency aim of 40 to 50 for 30-second counting time) when students did not make adequate progress with the 1-minute counting time. Students' Standard Celeration Charts, consulted by the teacher, helped guide the decision whether to make a change in timing length.

Passage fluency. Students practiced repeated readings of a passage they had read in the *Reading Mastery* program. The students in third and fourth grade practiced reading a passage until they met the Precision Teaching fluency aim of 200 words correct per minute (Beck, Conrad, & Anderson, 1995; Freeman & Haughton, 1993; Kubina, Amato, Schwilk, & Therrien, 2008). After a student met the fluency aim, he or she started to read a new passage and would do so again until reaching the aim. Students in second grade used a 30-second counting time and had an aim of 100 words. Students in first grade performed the repeated reading of the passage until they met an aim of 66 words in 20 seconds. If students could not read a minimum of 10 words in 30 seconds they did not engage in repeated reading.

The teacher selected stories for repeated reading. Passages came from a *Reading Mastery* passage that the students had already read. To implement the procedure, the teacher put the students into pairs with one student as the reader and the other as the scorer. Each student had a copy of the passage. The scorer placed a Mylar sheet over the passage. Once the teacher started the timer, students started to read while their partners used a dry-erase marker to write Xs by words the readers omitted or said incorrectly. At the end of the timing, the scorers shared feedback with the readers, wrote the scores on a separate datasheet, and then switched roles.

Each reader engaged in a repeated reading of the passage two to three times at the teachers' discretion. The teachers made their decision for the third extra practice trial based on the

trend of the data displayed on the Standard Celeration Charts. Teachers' decisions were influenced by flat or slowly growing trends in the data (cf. Figure 1). The teachers systematically checked partners' scoring accuracy by moving from student to student during each timing and varying their checking procedure each day to ensure they had an opportunity to observe all students.

Standard Celeration charting. Each teacher taught her class of students how to use the Standard Celeration Chart using a modified version of procedures described by Cancio and Maloney (1994). The Standard Celeration Chart procedures were found in a script that sequentially taught students to find day lines and counting lines and to display dots and Xs for correct and incorrect data. In first grade, approximately half of the students did not learn how to chart. Those students had either a classmate or paraprofessional help them. The teacher observed the charted frequencies and made decisions if a change to the particular practice procedure was warranted. Students could also participate in asking for a change or using a procedure they suggested (e.g., beating a set score for the day).

Results

Over the period of the six-week summer intervention program, both the celerations of students' learning and the standardized tests significantly increased (we report only the latter). Students showed statistically significant improvement from the pretest to posttest assessments for the informal and formal reading measures at the end of the six-week summer school program. Students who attended fewer than 25% of the summer school sessions were not included in the data analysis.

Informal measures

The changes in student learning are shown by a pretest and posttest for each measure using SPSS version 12 repeated measures analysis of variance (ANOVA) program, as indicated in Table 1. The first informal measure is the

Figure 1

The Decision Rules Chart Used By Teachers

Standard Celeration Chart data	Action
Meets aim for two out of three days	Make a change
Four to five days of flat data	Make a change
Minimum celeration less than x1.25 (for acceleration aims)	Make a change
Acceleration data decelerating	Make a change
Deceleration data accelerating	Make a change
Data fall below projected celeration aim line	Make a change
Teacher Prerogative (Teacher has information pertinent to improving the learner's performance)	Make a change

Adapted from Cancio & Maloney (1994) and other sources

number of *Reading Mastery* letter sounds said by a student in 1 minute. The frequency scores (n = 165) had a pretest mean of 42.4 (SD = 15.8) and a posttest mean of 64.2 (SD = 24.98). The improvement of 21.8 letter sounds per minute was statistically significant, F (1,164) = 173.035, p < .0005 and ? = .513, a moderate effect size (Vasquez, Gangstead, & Henson, 2000).

The second informal measure recorded one point for each correctly identified letter sound and for each word correctly read the fast way (n = 162). There was a statistically significant improvement from the pretest mean of 58.38 (SD = 24.26) to the posttest mean of 103.47 (SD = 42.86), a difference of 45.09 letters sounded out and words read per minute, F(1,161) = 241.207, p < .0005, ? = 0.60 (moderate effect size). The gain of 45 letters sounded out translates into an average gain of 9 to 11 more words orally decoded on a word list.

The third informal measure was the number of words read correctly (n = 148). There was also a statistically significant improvement from the pretest mean of 69.31 (SD = 32.29) to the posttest mean of 86.15 (SD = 40.96), an increase of 16.84, F(1,147) = 98.368, p < .0005, ? = 0.401 (moderate effect size). Students showed an average gain of 17 words per minute for their oral reading fluency.

Table 2 shows there were also statistically significant improvements on the selected standardized subtests of the *Woodcock Reading Mastery Test-Revised-NU* (Woodcock, 1998) for Word Identification (n = 97). The pretest mean was 89.04 (SD = 12.55) and the posttest mean was 94.00 (SD = 12.74). The difference between the means = 4.96, which was found to be a statistically significant improvement, F(1,96) = 20.741, p < .0005, with a small effect size of ? =.178.

In the second formal measure, Word Attack (n = 97), there was a pretest mean of 92.63 (SD = 18.28) and a posttest mean of 101.53 (SD = 14.94). The resulting difference between the means was 8.9, a statistically significant improvement, F(1,96) = 17.972, p < .0005, and a small effect size (? = 0.158).

Passage Comprehension (n = 93) was the final formal measure. Again, the difference between the pretest mean of 88.11 (SD = 14.11) and

Table 1

Pretest and Posttest Informal Measures of Reading Fluency

Informal reading measure	*n*	Pretest fluency mean	Posttest fluency mean	F	Effect size
Letter sound (identification) fluency	165	42.40 (SD= 15.80)	64.20 (SD= 24.98)	173.035**	0.513
Orally decoding words fluency	162	58.38 (SD= 24.26)	103.47 (SD= 42.86)	241.207**	0.6
Passage fluency	148	69.31 (SD= 32.29)	86.15 (SD= 40.96)	98.368**	0.401

** p < .0005

the posttest mean of 94.63 (SD = 12.37), a difference of 6.52, was statistically significant, $F(1, 92) = 30.220$, $p < .0005$, with a small effect size of ? = 0.247. Because of the moderate rather than large sample size and resulting empty cells, we did not separate the data according to the levels of *Reading Mastery* used for instruction. A larger sample size would have allowed the pretest-to-posttest changes in reading fluency to be evaluated in relationship to "in-program" *Reading Mastery* reading fluency goals.

Discussion

The combination of the DI program *Reading Mastery* and Precision Teaching implemented over the six-week summer school program resulted in statistically and educationally significant improvements in students' informal and formal measures of reading. In this intervention, summer school students received instruction from *Reading Mastery* programs and spent time practicing letter sounds, sounding out and saying words fast, and repeatedly reading passages to Precision Teaching fluency aims. The data are encouraging because they show that even over a short six-week summer school period, the reading skills of students greatly improved after being exposed to the combination of *Reading Mastery* and Precision Teaching.

This study supports the notion that Precision Teaching, combined with other curricula, produces positive outcomes (Lindsley, 1992). During the summer school implementation, teachers who used *Reading Mastery* continued to use the program as designed and did not change any formats or instructional delivery techniques. The addition of Precision Teaching required students to practice skills to fluency and to display data on a Standard Celeration Chart. The skills selected for the students to practice and monitor (letter sounds, sounding out words and saying them fast, and passage reading) were chosen because they are pivotal decoding skills. The scope and sequence for *Reading Mastery I, II,* and *III* all show that the selected skills used in this study play critical roles not only for decoding but also for comprehension. For example, oral reading fluency strongly reflects a student's overall reading competence (Fuchs, Fuchs, Hosp, & Jenkins, 2001).

Table 2

Pretest and Posttest Formal Measures of Reading Fluency

Formal reading measure	*n*	Pretest standard score mean	Posttest standard score mean	F	Effect size
Word Identification	97	89.04 (SD= 12.55)	94.00 (SD= 12.74)	20.741**	0.178
Word Attack	97	92.63 (SD= 18.28)	101.53 (SD= 14.94)	17.972**	0.158
Passage Comprehension	93	88.11 (SD= 14.11)	94.63 (SD= 12.37)	30.220**	0.247

** $p < .0005$

By facilitating fluency with pivotal decoding skills, beyond what the *Reading Mastery* program calls for, the use of Precision Teaching (i.e., practice methods, monitoring data on Standard Celeration Charts) may have produced a critical learning outcome associated with fluency called "application." Application refers to the process where component skills, when fluent, quickly apply or combine to form a composite skill (Binder, 1996; Haughton, 1972; Kubina & Morrison, 2000; Kubina, Young, & Kilwein, 2004). Some students, for instance, received instruction on letter-sound identification through *Reading Mastery I*. Letter-sound identification is a component skill of sounding out words. Students who could fluently identify letter sounds may have more readily applied the component skill to the composite behavior of sounding out a word more quickly than students who could not fluently say letter sounds. For example, students who could identify letter sounds at 100% accuracy but did so at a rate of 5 letter sounds per 10 seconds demonstrated a different performance sounding out words than students who identified 16 letter sounds per 10 seconds. Additionally, students who could sound out words fluently (i.e., 80 letters and words said fast per minute) may have applied this skill to the composite behavior of reading words in a passage more readily than students who orally decoded words at a rate of 30 letter sounds and words said fast per minute.

Because the teachers used Standard Celeration Charts to make instructional decisions, one would expect larger effect sizes for informal reading measures (i.e., directly practiced pivotal reading skills) than for formal measures (i.e., not directly practiced reading skills). The data show a larger effect size for the charted behaviors. The teachers looked at the fluency data on a daily basis and made instructional decisions following decision rules adapted from Cancio and Maloney (1994) and other sources (Figure 1). One decision rule—"If four to five days of flat data, make a change"—promoted active involvement and individualized and responsive changes implemented by the teacher. For instance, if a student did not make progress for three days for her letter sounds, the teacher analyzed the charted data, implemented an intervention, and then examined the results of the intervention in the coming days. Examples of interventions consisted of reducing the counting time or practice interval, having the student set goals, and selecting a school supply reward for obtaining an improvement goal (e.g., receiving a pencil after reading 15 more words correctly in 30 seconds).

The conclusions of this study present positive results, but there are several limitations that suggest alternative explanations. The methodology is also limited due to the one-group pretest-posttest design, but this method also allows for the investigation of the subject matter, which might otherwise not be feasible. In addition, the one-group pretest-posttest design includes variables such as history and maturation that can affect internal validity. Despite these limitations, readers are encouraged not to discount the findings of this study but instead to carefully interpret them.

This investigation cannot fully conclude that Precision Teaching augmented and improved the use of the DI program *Reading Mastery*. However, it can suggest that the effects of the combination of DI and Precision Teaching are positive for those students involved. Due to the lack of a control group and the study design, a cause-effect conclusion cannot be made, but the evidence does show a positive effect when using Precision Teaching with *Reading Mastery*.

Future Research

We hope to replicate the findings of the present study but with two additional control groups. Participants in the first control group would be pretested and posttested but would not participate in the intervention. Rather, they would receive alternative instruction that

did not include DI. This would allow us to see if the change in scores was due to our intervention rather than just an effect of going to summer school. A second control group would have a DI-only intervention, permitting an appraisal as to the effects of adding Precision Teaching to DI and using DI alone. Examining the separate effects for various interventions in the future holds value. For instance, do the Precision Teaching fluency aims for a particular skill enhance the progress students make in a specific strand (e.g., orally decoding words)? Also, additional research should be conducted to further confirm and establish the fluency aims (e.g., orally decoding words). Another suggestion for future research entails the analysis of disaggregated data by separate grades and reading levels. Many other future research questions may arise, and if the present study serves as an indication for prospective research, future students will benefit from the use of DI and Precision Teaching.

References

Adams, G. L., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR.* Seattle, WA: Educational Achievement Systems.

Beck, R., & Clement, R. (1991). The Great Falls Precision Teaching Project: An historical examination. *Journal of Precision Teaching, 8*(2), 8-12.

Beck, R., Conrad, D., & Anderson, P. (1995). *Basic skill builders handbook.* Longmont, CO: Sopris West.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*, 163-197.

Binder, C., & Watkins, C. L. (1989). Promoting effective instructional methods: Solutions to America's educational crisis. *Future Choices, 1*(3), 33-39.

Binder, C., & Watkins, C. L. (1990). Precision Teaching and Direct Instruction: Measurably superior instructional technology in schools. *Performance Improvement Quarterly, 3*(4), 74-96.

Blackwell, A., Stookey, S., & McLaughlin, T. F. (1996). The effects of using Direct Instruction and a re-reading contingency with Precision Teaching. *Journal of Precision Teaching and Celeration, 13*(2), 19-22.

Cancio, E. J., & Maloney, M. (1994). Teaching students how to proficiently utilize the Standard Celeration Chart. *Journal of Precision Teaching, 12*(1), 15-45.

Carnine, C. W., Silbert, J., Kame'enui, E. J., & Tarver, S. G. (2004). *Direct Instruction reading* (4th ed.). Upper Saddle River, NJ: Prentice Hall/Merrill.

Desjardins, E. A., & Slocum, T. A. (1993). Integrating Precision Teaching with Direct Instruction. *Journal of Precision Teaching, 10*(2), 20-24.

Edmonson, A., Peck, S. M., & McLaughlin, T. F. (1996). The effects of Direct Instruction on early reading skills of a kindergarten student. *Journal of Precision Teaching and Celeration, 14*(1), 72-76.

Engelmann, S., & Bruner, E. C. (1995a). *Reading Mastery I: Rainbow Edition.* Columbus, OH: SRA.

Engelmann, S., & Bruner, E. C. (1995b). *Reading Mastery II: Rainbow Edition.* Columbus, OH: SRA.

Engelmann, S., & Bruner, E. C. (1995c). *Reading Mastery I: Rainbow Edition — Teachers guide.* Columbus, OH: SRA.

Engelmann, S., & Hanner, S. (1995). *Reading Mastery III: Rainbow Edition.* Columbus, OH: SRA.

Fraenkel, J. R., & Wallen, N. E. (1996). *How to design and evaluate research in education* (3rd ed). New York: McGraw-Hill.

Freeman, G., & Haughton, E. (1993). Building reading fluency across the curriculum. *Journal of Precision Teaching, 10*(2), 29-30.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*, 199-208.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239-256.

Haring, N. G., & Krug, D. A. (1975). Evaluation of a program of systematic instructional procedures for extremely poor retarded children. *American Journal on Mental Retardation, 79*, 627-631.

Haughton, E. C. (1972). Aims: Growing and sharing. In J. B. Jordan & L. S. Robbins (Eds.), *Let's try doing something else kind of thing* (pp. 20-39). Arlington, VA: Council for Exceptional Children.

Hempenstall, K. (2004). The importance of effective instruction. In N. Marchand-Martella, T. Slocum, & R. Martella (Eds.), *Introduction to Direct Instruction* (pp. 1-27). Boston: Allyn & Bacon.

Holz, K. R., Peck, S. M., McLaughlin, T. F., & Stookey, S. (1996). The effects of using Direct Instruction and a re-reading contingency, coupled with a reward and praise contingency, with a high school sophomore. *Journal of Precision Teaching and Celeration, 14*(1), 35-40.

Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist, 47*, 1475-1490.

Johnson, K. R., & Layng, T. V. J. (1994). The Morningside model of generative instruction. In R. Gardner, D. Sainato, J. Cooper, T. Heron, W. Heward, J. Eshleman, & T. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 173-197). Belmont, CA: Brooks-Cole.

Johnson, K. R., & Street, E. M. (2004). *The Morningside model of generative instruction: What it means to leave no child behind*. Concord, MA: Cambridge Center for Behavioral Studies.

Kinder, D., Kubina, R., & Marchand-Martella, N. E. (2005). Special education and Direct Instruction: An effective combination. *Journal of Direct Instruction, 5*, 1-36.

Koorland, M. A., Keel, M. C., & Ueberhorst, P. (1990). Setting aims for precision learning. *Teaching Exceptional Children, 22*(3), 64-66.

Kubina, R. M. (2003). [College students' fluency sample performances of sounding out words and saying them fast]. Unpublished raw data.

Kubina, R. M. (2005). Developing reading fluency through a systematic practice procedure. *Reading & Writing Quarterly, 21*, 185-192.

Kubina, R. M., Amato, J., Schwilk, C. L., & Therrien, W. J. (2008). Comparing performance standards on the retention of words read correctly per minute. *Journal of Behavioral Education, 17*, 328-338.

Kubina, R. M., & Morrison, R. (2000). Fluency education. *Behavior and Social Issues, 10*, 83-99.

Kubina, R. M., Ward, M., & Mozzoni, M. P. (2000). Helping one person at a time: Precision Teaching and traumatic brain injury rehabilitation. *Behavioral Interventions, 15*, 189-203.

Kubina, R. M., Young, A. E., & Kilwein, M. (2004). Examining an effect of fluency: Application of oral word segmentation and letters sounds for spelling. *Learning Disabilities: A Multidisciplinary Journal, 13*, 17-23.

Lindsley, O. R. (1990). Precision Teaching: By teachers for children. *Teaching Exceptional Children, 22*(3), 10-15.

Lindsley, O. R. (1992). Precision Teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25*, 51-57.

Lindsley, O. R. (1997). Precise instructional design: Guidelines from Precision Teaching. In C. R. Dills & A. J. Romiszowski (Eds.), *Instructional development paradigms* (pp. 537-554). Englewood Cliffs, NJ: Educational Technology Publications.

Maloney, M. (1998). *Teach your children well: A solution to some of North America's educational problems*. Cambridge, MA: Cambridge Center for Behavioral Studies.

Marchand-Martella, N. E., Slocum, T. A., & Martella, R. C. (2004). *Introduction to Direct Instruction*. Boston: Allyn & Bacon.

Mercer, C. D., Mercer, A. R., & Evans, S. (1982). The use of frequency in establishing instructional aims. *Journal of Precision Teaching, 3*(3), 57-63.

Morrell, M., Morrell, R., & Kubina, R. M. (1995). Using Precision Teaching to enhance Direct Instruction sight-word reading. *The Journal of Precision Teaching and Celeration, 13*(1), 47-54.

National Research Council. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Neely, M. D. (1995). The multiple effects of whole language, Precision Teaching, and Direct Instruction on first-grade story-reading. *Effective School Practices, 14*(4), 33-42.

Pennsylvania Department of Education. *PSSA results: 2004-05 Performance Level Results by School District (and School) within Intermediate Unit/County*. Retrieved November 6, 2006, from http://www.pde.state.pa.us/a_and_t/site/default.asp

Stenseth, S., & McLaughlin, T. F. (1996). The effects of contingent consequences with Direct Instruction reading with a preschool child in the home. *The Journal of Precision Teaching and Celeration , 13*(2), 53-56.

Vasquez, L., Gangstead, S., & Henson, R. (2000, January). *Understanding and interpreting effect size measures in general linear model analyses*. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, TX.

West, R. P., & Young, K. R. (1992). Precision Teaching. In R. P. West & L. A. Hamerlynck (Eds.) *Designs for excellence in education: The legacy of B. F. Skinner* (pp. 113-146). Longmont, CO: Sopris West.

White, O. R. (1986). Precision Teaching—precision learning. *Exceptional Children, 52*, 522-534.

White, O. R. (2005). Precision Teaching. In M. Hersen, G. Sugai, & R. Horner (Eds.), *Encyclopedia of behavior modification and cognitive behavior therapy: Volume III. Education applications* (pp. 1433-1437). Thousand Oaks, CA: Sage.

Whitehurst, G. J. (2002). Evidence-based education. Paper presentation at the Student Achievement and School Accountability Conference. Retrieved April 2, 2004, from http://www.ed.gov/nclb/methods/whatworks/eb/edlite-slide001.html

Woodcock, R. W. (1998). *Woodcock Reading Mastery-Revised*. Circle Pines, MN: American Guidance Services.

MICHAEL R. VITALE, East Carolina University, and THEODORE S. KANIUKA, Fayetteville State University

Exploring Barriers to the Role of Corrective Reading in Systemic School Reform: Implications of a Three-Part Investigation

Abstract: In a three-part study we explored implications for enhancing the role of *Corrective Reading* (*CR*) in systemic school reform relating to teacher judgments of the high levels of student academic performance required by *CR* tasks in contrast to inappropriate instructional perspectives applied by teachers regarding possible adoption of *CR*. As an operational context, in Part 1 we demonstrated that a year-long implementation of *CR* in grades 3 and 4 with initially low-achieving, at-risk students significantly accelerated achievement growth in reading (vs. comparison students). In Part 2, we showed that teachers in demographically comparable non-*CR* schools judged content samples from *CR* as instructionally appropriate for higher-ability and more grade-advanced students than those with whom we had used the *CR* materials in Part 1. In Part 3 we found that, when asked to evaluate *CR* for possible adoption, teachers who were unfamiliar with *CR* held perspectives inconsistent with *CR* design. We present strategies for reconciling the findings across the three parts of the study and enhancing the perceived educational value of *CR* in school improvement.

Despite the limited decade-long improvements in reading in grades 4 and 8 reported by the National Assessment of Educational Progress (NAEP; Plisko, 2003), the absolute levels of student proficiency in recent NAEP assessments point to reading comprehension as a continuing systemic problem. The most recent 2005 NAEP assessment (NAEP, 2005) found that 27% of eighth graders could not read at the basic level, with only 31% of eighth graders performing at the proficient or advanced levels. The 2002 NAEP assessment (NAEP, 2002)—the most recent to include grade 12—found that 26% of 12th graders were not able to read at the basic level, and only 36% of 12th graders performed at the proficient or advanced levels.

Considering the implications of the preceding NAEP findings, large numbers of students are unprepared to extract the general meaning or make simple inferences from the texts they must read with understanding to be successful in school academic settings. Because such poor reading comprehension proficiency hinders student academic learning, it is not surprising that adolescents who are struggling readers are at risk of dropping out of high school without a diploma or graduating unprepared for college. Both circumstances are indicative of such students having limited opportunities in the workforce (Carnevale, 2001; National Center for Education Statistics, 2003).

As these recent NAEP findings have shown (e.g., Plisko, 2003), many students struggle

Journal of Direct Instruction, Vol. 9, No. 1, pp. 13-33. Address correspondence to Michael Vitale at vitalem@ecu.edu.

with basic reading skills such as decoding and word recognition. At the same time, many other readers who have gained basic reading skills at the elementary level also struggle with the vocabulary, fluency, and comprehension required for success in content-oriented instruction. With this point in mind, it is important that Plisko also called attention to a continuing and substantial achievement gap between white and African American students (30 points on the 500-point NAEP scale) and between students receiving free lunch services and students of low socioeconomic status (SES) who do not (22 points on the 500-point NAEP scale).

In fact, the recent RAND Reading Study (Snow, 2002) reported that the proficiency of students to comprehend subject-matter texts (i.e., content-area reading) has remained a significant problem in reading research. The RAND report noted that the reading achievement of 12th-grade students on NAEP's most recent (2002) long-term assessment represented a performance decrease from preceding years. Moreover, Biancarosa and Snow (2004) estimated that as many as 70% of older students are presently participating in school initiatives for the remediation of their deficiencies in reading skills. Although some significant research progress has been made on helping young children learn to read (e.g., Snow, 2002), very little research in the field of reading has determined how to remediate comprehension difficulties of students in content-oriented instruction in applied school settings through research-validated remedial reading practices (see Torgesen et al., 2006).

One important exception to the preceding reform scenario over the past 25 years has been *Corrective Reading* (*CR;* see Engelmann, Hanner, & Johnson, 1999), a Direct Instruction (DI) remedial reading program for students in grades 3–12 that has been extensively research-validated (see Grossen, 1998; Przychodzin-Havis et al., 2005). However, despite the substantial accumulation of scientific evidence supporting *CR* as an effective alternative for addressing the remediation of student reading deficiencies, *CR* has had minimal, if any, systemic impact on school improvement initiatives.

In an analysis of possible reasons for explaining the lack of prominence of *CR* in school reform, Kaniuka (1997) pointed to a variety of factors as potentially inhibiting the adoption of *CR*: (a) whether teachers who have an active role in the process of school reform have the foundational knowledge necessary for identifying sound instructional alternatives; (b) whether teachers' lowered academic expectations for at-risk, low-SES students influence their instructional decision-making; and (c) whether the established professional perspectives for "best practices" accepted in the discipline of education are counter to the design and implementation of *CR*. As Kaniuka reported, research (e.g., David, 1995-1996) has suggested that because of attention to the diverse and competing interests, much of the time and effort of many school-based improvement teams is devoted to issues that are unrelated to student achievement outcomes. To the extent that these dynamics dominate instructional decision-making within school improvement processes, *CR* is at a substantial disadvantage in being considered for adoption.

In the context of the preceding observations, the purpose of this three-part study was to investigate systemic implications for enhancing the role of *CR* in the dynamics of school reform. The two forms of implications addressed were complementary. They were (a) whether teachers' expectations substantially underestimated the potential levels of achievement in reading decoding and comprehension of their students relative to those achieved by remedial reading students in *CR* and (b) whether the prior professional perspectives of teachers involved in an instructional decision-making simulation regarding

the possible adoption (or non-adoption) of *CR* in their schools were inconsistent with *CR* design. Thus, this study explored whether the prior perspectives of teachers not familiar with the specific student academic outcomes resulting from *CR* would preclude the program being considered for adoption. The study also determined possible strategies that would support consideration of *CR* for use in school improvement initiatives.

Part 1 of this study provided a foundation by documenting positive achievement outcomes resulting from a year-long implementation of *CR* with low-proficient, remedial readers in grades 3 and 4 in a low-SES school. Using the results of Part 1 as a framework, Part 2 investigated the perceptions of nonprogram teachers (i.e., teachers in schools not using DI) in both SES-comparable (i.e., percentage of students receiving free or reduced-cost lunch) and more advantaged schools with regard to the expected performance of their students on representative decoding and comprehension content taught in *CR* during the year-long implementation in Part 1. Part 3 engaged a new group of nonprogram teachers in a school decision-making simulation in which they evaluated *CR* in a preliminary fashion and made recommendations regarding the possible consideration of *CR* for adoption by their school.

Considered together, the three parts of this study addressed a set of interdependent research questions regarding the potential role of *CR* as a DI program in school reform settings. The specific research questions were as follows:

1. Did the year-long implementation of *CR* in grades 3 and 4 result in greater overall student academic growth in reading by low-achieving, low-SES students in comparison to control students, as measured by a state accountability reading assessment test?

2. Did the academic expectations of nonprogram teachers of the possible performance of their students on representative content from *CR* mastered by the low-achieving, low-SES students in Part 1 substantially underestimate the achievement potential of students in their schools?

3. What were the recommendations of a new sample of nonprogram teachers regarding the potential adoption of *CR* in their schools obtained in a simulated decision-making activity, and to what degree were their prior instructional perspectives regarding remedial reading consistent with *CR* program design features?

In all three parts of this study, the students (Part 1) and the two different groups of teachers (Parts 2 and 3) were selected from demographically comparable elementary schools located in eastern North Carolina. All of the schools were located in comparable small towns in rural areas. All of the schools were racially mixed and had achievement levels that ranged from below average to slightly above average as measured by state-administered, accountability achievement tests. In keeping within a state-mandated educational reform initiative, all of the schools had been involved in school reform and instructional decision-making initiatives emphasizing basic skills improvement for at least two years at the time of this study.

Part 1: Effectiveness of CR *in an Eastern NC School*

The purpose of Part 1 was to determine whether *CR* was effective in accelerating student reading achievement within the Eastern North Carolina demographic setting.

Method

Participants. The school implementing *CR* was a historically low-achieving grade 3–5 school (N = 376, 73% minority), with reading scores on state reading tests well below district and

state averages. The participants assigned to *CR* were the lowest-performing 20% of grade 3 and grade 4 students at the beginning of the school year. The comparison students for reading achievement consisted of the lowest-achieving 20% of students in grade 3 and grade 4, respectively, enrolled in the same school the preceding school year. More specifically, the preceding year's grade 3 reading achievement of the grade 4 treatment group served as a comparison group for the grade 3 treatment students, and the grade 4 reading achievement of the school's grade 5 (nonparticipating) students served as a comparison group for the grade 4 treatment group. The use of the specific within-school comparison groups in this study allowed the achievement outcomes anticipated from *CR* to be interpreted in terms of potential value of *CR* for the school's low-SES student population, illustrating what the students' prior achievement could have been if they had been engaged in *CR* the year preceding this study.

Instruments. Student academic reading achievement was measured using the *North Carolina End-of-Grade Tests in Reading (EOG-R)* for grades 3 and 4 (North Carolina Division of Accountability Services, 2006). The *EOG-R* is a state-developed test reflecting North Carolina instructional standards (North Carolina Department of Public Instruction, 2004). Student *EOG-R* performance is reported using an equal-interval developmental scale score (DSS) across grades 3 through 8. At each grade level, state-established DSS cut-scores establish four levels of student reading proficiency, with Levels 1 and 2 indicating below grade-level performance and Levels 3 and 4 indicating at/or above grade-level performance. In this study, *EOG-R* scale scores in grades 3 and 4 were used as achievement outcome measures, with the preceding year's grade 3 *EOG-R* scores serving as a concomitant variable (covariate) for grade 4 students.

The *Cognitive Abilities Test* (*CogAT*; Lohman & Hagen, 2001) verbal subtest provided ability scores at the beginning of grade 3 and served as a covariate for grade 3 and grade 4 students. Because *EOG-R* tests are not administered in grade 2, the *CogAT* score was the sole covariate for the grade 3 students.

CR instruction. CR (see Engelmann, Hanner, & Johnson, 1999) was designed as a core replacement (vs. a supplementary) remedial reading program in which both decoding and comprehension skills and strategies are taught explicitly through direct teacher-led instruction in a manner that is suitable for students in grades 3-12. The *CR* programs used in this study (Engelmann, Carnine, & Johnson, 1999; Engelmann et al., 1999a, 1999b, 1999c, 1999d, 1999e) consisted of two complementary parts, *Decoding* (*CR-D)* and *Comprehension* (*CR-C*), which are comprised of a total of 320 and 325 lessons, respectively. *CR-D* consists of Levels A, B1, and B2 (65 lessons each) and C (125 lessons). *CR-C* consists of Levels A and B1 (60 lessons each), B2 (65 lessons each), and C (140 lessons).

CR placement tests determine which program level is appropriate for students. Small instructional groups ranging in size from 6 to 12 students are used to implement the program. Students initially assigned to instructional groups through placement tests are evaluated continually during daily lessons using program mastery learning criteria (e.g., error rates in oral components of lessons, performance on embedded lesson tests). Daily lessons for each strand require approximately 40-45 minutes to complete. As a replacement (vs. a supplementary) program, the 90 minutes of instructional time allocated to *CR* was equivalent to the amount of instructional time allocated to the district-adopted basal reading series received by the comparison groups used in the study.

In the present study, all 27 grade 3 treatment students and 25 of the 33 total of

grade 4 students placed at Level A in both *CR-D* and *CR-C*. The remaining 8 students in grade 4 placed at Level B1 in *CR-D* and Level A in *CR-C*. Using student placement test performance to group students for instruction resulted in classes ranging in size from 7 to 10 students.

The *CR* program (*CR-D, CR-C*) was implemented over the entire school year within daily lesson blocks lasting 90 minutes (i.e., 45 minutes for each strand). All 8 teachers received *CR* training on how to administer placement tests in the preceding spring, 4 hours of program-specific training prior to the start of the school year, and 3 additional follow-up sessions during the school year for a total of 10 hours. In addition, an experienced DI consultant and the principal observed all *CR* teachers regularly to evaluate teaching effectiveness. All *CR* teachers met as a team once every 2 weeks for 90 minutes to discuss program implementation and other curriculum-related issues.

In grade 3, all three groups of students placed in Level A of *CR-D* and Level A of *CR-C*. Of these grade 3 groups, for *CR-D*, one group completed Level B1, one completed Level B2, Lesson 10, and one completed Level B2, Lesson 20. For *CR-C*, one group completed Level B1, one completed Level B1, Lesson 10, and one completed Level B1, Lesson 60.

In grade 4, three of the four groups of students placed in Level A of *CR-D* and Level A of *CR-C*. The remaining group placed in Level B1 of *CR-D* and Level A of *CR-C*. Of these grade 4 groups, for *CR-D*, three of the four groups reached Level B2, with one group completing Lesson 5, one completing Lesson 10, and one Lesson 15. The fourth group completed Lesson 55 in Level B2. For *CR-C*, one group completed Level A, one group completed Level B1, Lesson 5, one completed Level B1, Lesson 50, and one completed Level B2, Lesson 10. The number of lessons completed ranged from 125 to 150 for *CR-D* and from 115 to 140 for *CR-C*.

Table 1

Demographic Characteristics of the Treatment and Comparison Groups for Academic Achievement in Reading

Grade	Group	N	Gender M	Gender F	Race B	Race H	Race W	Exceptionalities[a]
3	Treatment	20	9	11	18	2	0	6 (5 LD, 1 SI)
	Comparison[b]	22	14	10	21	0	0	7 (4 LD, 1 EMH, 1 EH, 11 H/SI)
4	Treatment	24	14	10	21	0	0	7 (4 LD, 1 EMH, 1 EH, 11 H/SI)
	Comparison	20	8	12	16	2	2	0

[a] LD = Learning Disabled, SI = Speech Impaired, EMH = Educable Mentally Handicapped, EH = Emotionally Handicapped, H/SI = Hearing/Speech Impaired.

[b] The grade 3 comparisons were the grade 4 treatment students the preceding year. All treatment and comparison groups consisted of the lowest-achieving students in their grade.

Design, analysis, and procedure. All students in the treatment and comparison groups met criteria indicative of low academic performance (below median *CogAT Verbal Subtest* scores in grade 3 and a Proficiency Level 1 on state-administered *EOG-R* tests in grade 3 indicated a substantial reading deficiency). Table 1 shows demographic characteristics of the treatment and comparison groups in Part 1 of the study.

Separate statistical analyses were conducted for the third- and fourth-grade samples because they had different covariates (*CogAT* in grade 3 versus *CogAT* and *EOG-R* in grade 4). The grade 3 ANCOVA compared the performance of the treatment and comparison groups on grade 3 *EOG-R* scores, using the *CogAT* verbal subtest as a covariate. A similar ANCOVA was conducted for grade 4 students on *EOG-R* performance, but with both the grade 3 *EOG-R* and *CogAT* verbal subtest scores used as covariates. Prior to conducting the ANCOVA for both grades, an independent sample *t*-test comparing the treatment and comparison groups' covariate scores was conducted to ensure that the initial levels of performance by the two groups on the concomitant variables were comparable.

All student *EOG-R* achievement tests were administered in late Spring through the North Carolina standardized testing program. Student *CogAT* verbal subtests were administered to all grade 3 students at the beginning of the school year through a local district-wide testing program. All resulting achievement data (*EOG-R, CogAT*) were obtained electronically from district records.

Results

Grade 3 findings. Table 2 shows the means and standard deviations for the grade 3 treatment and comparison students on the *CogAT* Verbal Subtest scores and the *EOG-R*. An independent *t*-test found no significant difference in the initial levels of performance between the treatment and comparison groups, $t\ (40) = 0.213$, $p = .83$, in verbal ability as measured by the *CogAT* at the beginning of grade 3. Thus, the treatment and comparison groups in grade 3 were comparable in prior academic achievement.

The results of the analysis of covariance (ANCOVA) comparing the adjusted mean *EOG-R* reading achievement of the treatment and comparison groups found the *EOG-R* achievement of the *CR* students significantly greater than comparisons, $F\ (1, 39) = 5.43$, $p < .05$. This finding consisted of a difference between adjusted means of 3.1 scale points on the *EOG-R* achievement test in favor of the treatment group, approximately 24% more than one year's *EOG-R* achievement growth, and an effect size of .72 (Cohen, 1988).

Grade 4 findings. Table 3 shows the means and standard deviations for the grade 4 treatment and comparison students on the grade 3 CogAT verbal subtest and the grade 3 *EOG-R* test. An independent *t*-test conducted on the *CogAT* verbal subtest scores, $t\ (42) = -0.052$,

Table 2

Performance of Grade 3 Treatment and Comparison Students on the CogAT Verbal Subtest and EOG-R

Group	*n*	*CogAT* Verbal	*EOG-R*
Treatment	20		
Mean		73.3	131.3
SD		9.8	6.1
Comparison	22		
Mean		72.7	128.0
SD		9.0	3.5

Note. CogAT scores are percentile ranks, *EOG-R* scores are developmental scale scores.

$p = .96$ found no statistically significant difference between the *CogAT* verbal subtest scores for the grade 4 treatment and comparison groups at the beginning of grade 3. A second independent *t*-test conducted on the *EOG-R* scores, $t\ (42) = 1.729, p = .09$, also indicated no statistically significant difference between the grade 4 treatment and comparison groups at the end of grade 3. Thus, the treatment and comparison groups in grade 4 were considered comparable in prior academic achievement.

The results of the ANCOVA comparing the adjusted mean *EOG-R* achievement of the grade 4 treatment and comparison groups found the reading achievement of the *CR* students significantly greater than comparisons, $F\ (1, 40) = 7.30, p < .03$. This finding consisted of a difference between adjusted means of 2.8 scale points on the *EOG-R* achievement test in favor of the treatment group, approximately 29% more than one year's *EOG-R* achievement growth, and an effect size of .88 (Cohen, 1988).

Summary and Discussion of the Results for Part 1

Part 1 found significant differences in favor of both the grade 3 and grade 4 *CR* students on reading achievement (*EOG-R*) relative to comparison students and showed that *CR* accelerated the rate of student growth by 24% in grade 3 and 29% in grade 4. These findings demonstrated that in the eastern North Carolina setting, *CR* was an effective intervention for accelerating student reading achievement as measured by state accountability tests.

Part 2: Expectations of Nonprogram Teachers of the Performance of their Students on CR *Content Samples*

Using Part 1 as a foundation, Part 2 of this study explored (a) the curricular judgments of nonprogram teachers in schools not using *CR* on the academic level of representative *CR*

Table 3

Performance of Grade 4 Treatment and Comparison Students on the CogAT Verbal Subtest and EOG-R in Grade 3 and the EOG-R in Grade 4

Group	*n*	Grade 3		Grade 4
		CogAT Verbal	EOG-R	*EOG-R*
Treatment	24			
Mean		77.5	128.0	138.3
SD		6.5	3.8	3.7
Comparison	20			
Mean		75.4	131.1	135.3
SD		6.3	6.5	3.5

Note. CogAT scores are percentile ranks, EOG-R scores are developmental scale scores.

content in terms of grade-level appropriateness for their low-, average-, and high-ability students and (b) the percentages of students in their school and district that nonprogram teachers estimated would demonstrate mastery of the *CR* content sampled. Together, the curricular judgments by nonprogram teachers in Part 2 addressed the question of whether *CR* as a remedial instructional intervention would be revealed by teacher expectations to have the potential to accelerate the rate and quality of student learning, an important issue for adopting *CR* in school reform.

Method

Participants. Fifty-seven nonprogram teachers from six K-5 elementary schools not using *CR* (or *DI*) participated in this part of this study. Because the purpose of Part 2 of this study was to obtain teacher judgments of the performance of their students on *CR* instructional activities, the school sample encompassed a range of performance levels from below average (approximately matching the school implementing the program in Part 1) to slightly above average. Specifically, the prior *EOG Reading* achievement across grades 3 and 4 for two of the five participating schools was less that 6 scale points higher than the experimental school in Part 1, but three of the five participating schools scored from 8 to 17 points higher. Of the five schools, two had fewer (34%, 65%) and three had greater (94%, 95%, 96%) enrollments of low-SES students than the school in Part 1 (71%). The group of five participating schools was comparable to the Part 1 school in key teacher demographics (gender, race, education, and experience).

Instruments. The *Student Activity Analysis Form* (*SAAF*; Vitale, Boldt, Kaniuka, & Scott, 1999) was used to assess teacher judgments regarding the grade-level appropriateness and degree of student mastery on 10 representative student *CR* activities (i.e., instructional tasks) selected from *CR-D* and *CR-C*. The activities were selected only from Levels A, B1, and B2 (Engelmann, Carnine, & Johnson, 1999; Engelmann et al., 1999a, 1999b, 1999c, 1999d, 1999e) rather than from the more advanced Level C, because Levels A, B1, and B2 represented the range of *CR* content taught across grades 3 and 4 at the treatment school in Part 1. For each task on the *SAAF* (e.g., reading passages, inference tasks), teachers were asked to make two different judgments: (a) the grade level at which the task would be appropriate for low-, average-, and high-ability students in their school and (b) the percentage of students in their school and in their district that would display mastery of these activities if tested.

The program locations of the 10 randomly sequenced *CR* activities consisted of two *CR-D* tasks (Level B1–Lesson 60 and Level B2–Lesson 30) and eight *CR-C* tasks (Level A–Lesson 50, Level B1–Lessons 19, 38, 39, 53, and 60, and Level B2–Lessons 32 and 43). Only two *CR-D* tasks were included because they adequately represented the major focus in Level B1 and B2 (*CR-C* tasks were far more diverse). The two *CR-D* tasks each consisted of 100-word passages that students in CR are required to read with fluency while meeting program-specified rate and accuracy criteria. The eight *CR-C* tasks included examples of rule statements, deductions, analogies, vocabulary, sentence parts, and contradictions on which students in *CR* are required to demonstrate mastery.

Although teachers responded anonymously, the *SAAF* scale included a cover page on which teachers supplied general demographic information (e.g., gender, grade taught, overall teaching experience, education level, and race). The next two *SAAF* pages consisted of illustrative (non-*CR*) activities designed to orient the participants to the *SAAF* response format. The next 10 pages presented the representative activities (one per page) from the *CR-D* and *CR-C* lessons.

In rating each *CR* task/activity, participants were asked to follow a two-step process. First, teachers were asked to estimate the grade level at which each task would be most appropriate instructionally for students they would consider to be of average, low, and high ability in their school. Second, teachers were asked to estimate the percentage of students in their school and district that, for each grade from grades 3 through 7, they believed would display mastery on each task (if tested). For grades 6 and 7, teachers were asked to estimate likely mastery based on their experience with past students. This was considered appropriate because the emphasis of the *SAAF* task was to assess teacher perceptual judgments about *CR* tasks rather than to provide absolute judgments of student performance. In responding, participants were free to assign the grade levels and percentages they deemed appropriate. Prior to use in this study, the *SAAF* format and instructions were pilot-tested and revised as necessary. In no case were teachers informed that the tasks to be judged came from *CR*.

Design, analysis, and procedure. In surveying nonprogram teachers, principals at each sample school were contacted in the third week in April of the school year to review *SAAF* administration procedures and materials. In each school, the principals subsequently distributed the *SAAF* to all of their teachers in a faculty meeting after explaining the purpose of the survey task. After completing the two practice pages, principals guided teacher inspection of the 10 *SAAF* tasks. Teachers then either completed the *SAAF* during the meeting or returned their completed forms to the principal in one week. In all cases, teachers were asked to make their judgments independently of other teachers.

The teacher grade-level expectations were analyzed using a 2 x 3 repeated measures ANOVA with two within factors: Program Content (*CR-D, CR-C*) and Student Ability (Low, Average, High). The teacher estimates of student mastery were analyzed using a 2 x 2 x 5 repeated measures ANOVA with three within factors: Program Content (*CR-D, CR-C*), Location (District, School), and Grade (3, 4, 5, 6, 7).

Results

Perceived grade-level assignment for CR tasks. Table 4 shows the means and standard deviations for the appropriate grade levels assigned to the *CR* tasks for low-, average-, and high-ability students. As Table 4 shows, the grade levels assigned by nonprogram teachers for the *CR-D* activities ranged from 3.3 for high-ability students to a high of 5.4 for low-ability students, with an average grade level for all students of 4.3. The grade levels assigned the *CR-C* tasks ranged from 4.1 for high-ability students to a high of 5.7 for low-ability students, with an average grade level for all students of 5.2. Thus, the *CR-D* and *CR-C* tasks completed by the low-SES remedial students in grades 3 and 4 in Part 1 were rated by teachers as appropriate for elementary students at much higher grade levels, with only high-ability students

Table 4

Nonprogram Teacher SAAF Scores for Grade-level Appropriateness as a Function of CR Program Content and Student Ability

Student Ability		*CR Decoding*	*CR Comprehension*
Low	*m*	5.41	5.69
	sd	1.09	.93
Average	*m*	4.26	5.69
	sd	.85	.96
High	*m*	3.26	4.14
	sd	.85	.86

expected to master the specified tasks at or below the fourth-grade level.

The results of the 2 x 3 ANOVA with the two factors, Program Content (*CR-D, CR-C*) and Student Ability (Low, Average, High) using teacher-estimated grade-level appropriateness as the dependent measure, found both the main effects and interaction significant. The significant Program Content effect, F (1,56) = 60.72, $p < .001$, indicated that teachers viewed the *CR-C* tasks included in the *SAAF* as more difficult than the *CR-D* tasks across all ability levels. The significant Student Ability effect, F (2,112) = 362.44, $p < .001$, confirmed that teachers were consistent in rating the grade-level appropriateness of the tasks lower (i.e., appropriate at a lower grade level) for high-ability students, next lowest for average-ability students, and highest (i.e., most difficult) for low-ability students, and that teachers expected different ability students to master the tasks at different grade levels (i.e., high-ability students would be expected to master the tasks at a mean grade level of 3.7, average-ability students at a mean grade level of 5.0, and low-ability students at a grade level of 5.6).

In addition to the significant main effects, a significant Program Content x Student Ability interaction, F (2,112) = 136.21, $p < .001$, showed that the differences in grade-level expectations between student ability levels were different for the *CR-D* and *CR-C* content. For low-ability students, the differences in grade-level expectations between students of different abilities (average-low, high-average) were 1.1 and 1.0 for *CR-D*, but 0.0 and 1.6 for *CR-C*. To explore this effect further, a single degree of freedom polynomial contrast revealed a significant, F (1,56) = 193.93, $p <$ .001, interaction between the linear effects of Program Content x Student Ability.

Expected student performance on CR tasks. Table 5 presents descriptive statistics for the nonprogram teacher expectations of the percentage of students in grades 3-7 in their schools and districts who would be able to demonstrate mastery of the 10 *CR-D* or *CR-C* tasks sampled from *CR*. As Table 5 shows, teachers rated the *CR* tasks as difficult for the majority of their students in Grades 3 and 4, with the findings showing an expected positive relationship between grade level and percentage of students capable

Table 5

Nonprogram Teachers' SAAF Estimates of the Percentage of Students Mastering CR-D *and* CR-C *Tasks as a Function of Grade Level and Location (School, District)*

		Program Content									
Student		Grade Level–*CR Decoding*					Grade Level–*CR Comprehension*				
Group		3	4	5	6	7	3	4	5	6	7
School	*m*	31.5	46.9	59.3	69.0	77.7	22.5	35.4	48.9	61.3	71.3
	sd	22.2	24.1	24.4	24.8	22.5	14.5	18.7	21.1	22.1	22.7
District	*m*	33.4	47.5	62.1	71.5	79.7	23.9	37.7	52.9	65.1	74.4
	sd	23.0	25.5	23.5	22.9	19.9	14.7	18.2	21.3	20.8	20.6
Total	*m*	32.6	46.7	60.7	70.1	78.7	23.2	36.5	50.9	63.2	72.9

Note. $N = 57$.

of mastering the tasks. Although the rate of increase was not uniform across the grades, it was consistent at the district and school levels. In addition, supporting the anticipated difficulty of the *CR* tasks, the expected mastery-level percentages for students were below 70% in all but five cases. Across grade levels, the expectations for success on *CR-D* tasks ranged from a low of 32.5% to a high of 77.8%, for a difference of 45.3%. On *CR-C* tasks, the expectations ranged from a low of 23.2% to a high of 72.9%, for a difference of 49.7%.

The results of a 2 x 2 x 5 repeated measures ANOVA with three within factors, Program Content (*CR-D, CR-C*), Location (District, School), and Grade (3, 4, 5, 6, 7) using teacher-estimated task mastery of students as the dependent variable found no significant difference in judgments between Location (School versus District), F (1,56) = 3.11, $p > .05$, indicating that teachers rated the overall mastery of students in their schools to be similar to those in their districts. At the same time, the significant Grade effect, F (4,224) = 450.02, $p < .001$, confirmed that teachers expected higher levels of mastery from students in higher grades. Together these findings confirm the general credibility of the teacher judgments. The ANOVA also found a significant Program Content effect, F (1,56) = 41.25, $p < .001$, a finding that the *CR-D* tasks were viewed as less difficult than the *CR-C* tasks (mean difference = 8.4%).

Among the interactions, only the Program Content x Grade was found significant, F (4,224) = 3.74, $p < .02$. This showed that the differences between the mean mastery percentages for *CR-D* and *CR-C* varied differentially across grade levels.

Summary and Discussion of the Results for Part 2

Part 2 of this study showed that nonprogram teachers having no prior experience with *CR* rated the *CR-D* and *CR-C* tasks as being composed of "grade-advanced" content, certainly well beyond the expected mastery of the majority of the third and fourth graders in their non-*CR* schools, in general, and low-achieving third and fourth graders, in particular. Further, the nonprogram teacher estimates of the percentages of students in their schools and districts who could be expected to display mastery revealed the *CR* tasks to be highly difficult, with mastery percentages for all grades being well below what would normally be considered acceptable mastery percentages of student performance (e.g., 80% mastery). Finally, in focusing upon the grade-level appropriateness of the *CR* tasks, teachers clearly viewed them as being more appropriate for average- and high-ability students at higher grade levels than the remedial grade 3 and 4 students who demonstrated mastery of them through *CR* in Part 1. In a similar fashion, the nonprogram teacher estimates of mastery of the *CR* tasks by their students showed that the adoption of *CR* would have substantially accelerated the rate and level of academic achievement for grade 3 and 4 students in their schools and districts.

In considering these nonprogram teacher estimates further, it also is important to remember that only representative tasks from Levels A, B1, and B2 of *CR* were sampled (with only two selections included from *CR-C* Level B2) because the intent was to stay within the scope of the *CR* Levels experienced by the grade 3 and 4 remedial students in the treatment school in Part 1. As a result, it is a reasonable expectation that selections from the more advanced Level C of *CR-D* and *CR-C* would be rated by teachers as even more advanced instructionally. Again, teacher perspectives on the degree to which *CR* could be expected to accelerate student learning should be an important issue in school reform initiatives focusing on student achievement.

Part 3: Analysis of Teacher Instructional Perspectives in a Simulated Decision-making Activity Evaluating CR *for Adoption*

Part 1 demonstrated that *CR* significantly accelerated the reading achievement of previously low-achieving, low-SES students in grades 3 and 4. Part 2 showed that the perceptual judgments of nonprogram teachers on the performance of their students on the representative *CR* content mastered by grade 3 and grade 4 students in Part 1 established *CR* as engendering an accelerated rate of achievement growth. Combining Parts 1 and 2, an instructional implication is that the achievement of a majority of students in the Part 2 schools, including those more academically advanced and older, would benefit academically from Levels A, B1, and B2 of *CR*.

With Parts 1 and 2 as a foundation, Part 3 engaged a new group of teachers having no prior, direct, in-depth experience with *CR* (or *DI*) in a decision-making process designed to simulate a preliminary evaluation of *CR* for possible adoption as a remedial reading program for their school.

Method

Participants. Elementary teachers ($N = 21$) enrolled in a graduate elementary education course and representing 13 different eastern North Carolina schools participated in Part 3. All teacher participants were enrolled in a graduate course in elementary education. None of the teachers or schools participating in Part 2 participated in Part 3. Two of the schools in Part 3 had four teachers in this study, one school had three teachers, and the remainder had one participant. Eighteen participants held a bachelor's degree (with 13 enrolled in a masters program) and three held a master's degree in another area. Two participants were male and 19 were female. One participant was black and the remaining 20 participants were white. The average teaching experience was 6.9 years, with a range from a low of one year of experience to a high of 29 years of experience. The grade levels taught by the teachers ranged from kindergarten to grade 6. None of the teacher participants had any direct or detailed prior knowledge of *CR*, although one teacher had observed *CR* being taught in a special education setting in her school on an informal basis.

Instrument. A researcher-developed paper-pencil *Teacher Decision-making Simulation* (*TDS*) was used to structure individual teacher activity designed to simulate key elements of a school instructional decision-making process. In the *TDS* activity, teachers were asked to assume they were to develop preliminary recommendations to their school improvement team regarding possible adoption of *CR* for remedial reading instruction in their school. The *TDS* provided teachers with a short, 1-page overview of *CR* implementation features, a copy of the *CR Series Guide* (Engelmann, Hanner, & Johnson, 1999) to use, and instructions to complete a preliminary evaluation of *CR* in approximately 50 minutes. The *TDS* was field-tested informally and refined prior to use in the study.

The *TDS* implementation instructions asked the participants to provide demographic information and then to review the short, one-page list of key factual details on the one-page overview extracted from the *CR Series Guide* (e.g., the program included a *CR-D* and a *CR-C* component, there were three levels (A, B, C) in each, materials for each level consisted of teacher presentation books and student workbooks, recommended daily small-group instructional time was 45 minutes for *CR-D* and 45 minutes for *CR-C*). The providing of these factual details allowed teachers to focus their attention on the *CR* curricular content and teaching strategies rather than having to

determine specific implementation details. In responding to the *TDS*, teachers were asked to support their responses with professional knowledge as appropriate.

After providing demographic information and reviewing the factual program information, participants were asked to assume their goal was to make a preliminary recommendation regard-

Table 6

Six Sets of Questions That Guided Teachers in the Teacher Decision-making Simulation (TDS)

Question 1.	Briefly outline your general school achievement goals in reading and your judgment of the general progress of your school meeting these goals.
Question 2a.	Estimate what percentage of your school's population is in a remedial reading program.
Question 2b.	Using the following scale, in your opinion, how successful is your school's present remedial program in meeting the needs of your students? Very Successful Successful Somewhat successful Unsuccessful Very Unsuccessful
Question 2c.	What general observations can you offer about the remedial reading program in your school?
Question 3.	In evaluating *Corrective Reading*, what would you identify as its major strengths/best features (if any) as a remedial reading program? (In your evaluation comments please address the instructional content, teaching strategies, and implementation requirements.)
Question 4.	In evaluating *Corrective Reading*, what would you identify as its major weaknesses/worst features (if any) as a remedial reading program? (In your evaluation comments please address the instructional content, teaching strategies, and implementation requirements.)
Question 5.	Based on your evaluation of *Corrective Reading*, and your knowledge of your school's students and reading program(s), would you recommend further consideration of *Corrective Reading* for possible adoption? If you would not recommend *Corrective Reading*, please indicate your recommendation.
Question 6.	If you could design an "ideal" remedial reading program, what important features would it have?

ing further consideration for possible adoption of *CR* for remedial reading to their school improvement team. In completing the simulated evaluation of *CR*, teachers were asked to inspect the *CR Series Guide* and then respond to the open-ended questions located on separate pages of the *TDS* instrument summarized in Table 6. In the *TDS*, questions 1, 2a, 2b, and 2c provided a context for the simulation task, while questions 3, 4, 5, and 6 provided qualitative data regarding teacher recommendations and their associated perspectives as applied to the decision-making task.

Design, procedure, and data analysis. The participants completed the *TSD* during a regular class meeting of a graduate elementary education course. All participants responded individually to the *TDS* because the emphasis of the study was to capture their individual decision-making perspectives rather than to emulate a typical group-based school improvement committee process. In administering the *TSD*, the instructor read the task directions to the class and was present to answer any questions pertaining to the evaluation. Participants were given a packet containing the *TDS* materials, which included the 1-page overview and the *CR Series Guide*. As the participants explored the *CR Series Guide*, they wrote their answers to the questions on the *TDS* form and then returned their completed materials.

The resulting teacher *TDS* responses were categorized in terms of (a) recommendations regarding the preliminary adoption of the program, (b) the professional rationale used to justify recommendations, and (c) misinterpretation of key characteristics of *CR*. In particular, the analysis focused on Questions 3 and 4 on the *TDS* that asked participants to evaluate *CR* and reference specific aspects of the program that they felt were strengths or weaknesses. These responses were analyzed by a panel of raters to identify the criteria reported by the participants in evaluating the *CR* materials and in supporting their recommendation decisions rather than simply summarizing the participant knowledge about remedial reading instruction per se.

Question 5 of the *TDS* asked teachers to either recommend *CR* in a preliminary fashion or make an alternative recommendation based on their knowledge of the instructional needs in their schools. The resulting teacher recommendations were categorized into three response categories: (a) positive, (b) negative, and (c) negative with a qualifying condition. Because all three recommendations were accompanied with supporting details, the response categories used in Questions 3 and 4 were again applied in categorizing the data in Question 5. Finally, Question 6 responses were organized using a three-category classification similar to the preceding: (a) teaching strategies, (b) instructional content, and (c) assessment. Although the focus of the *TDS* was on teaching strategies and instructional content, the category of assessment was added because teacher beliefs about assessment and student achievement emerged as a factor regarding participant judgments of the potential effectiveness of remedial programs in their schools.

Compilation of the resulting qualitative data consisted of categorization and frequency counts by category by one of the researchers and by three other independent raters, all of whom were doctoral students in educational leadership. The percentage-of-agreement reliability coefficients indicating the degree of agreement in following the scoring systems among the four raters ranged from .88 to .62, with a mean of .75.

Results

School contexts for remedial reading. The percentage of students assigned to remedial reading programs in teachers' schools ranged from 5% to 100%, with a mean of 31%. The majority of the 20 responding teachers reported that their schools were either successful (20%) or somewhat successful (65%) in meeting the needs of their remedial students. Although teachers

were generally supportive of the efforts of their schools in remediating reading deficiencies, many stated that not enough students were being served in their schools.

Teacher recommendations for CR adoption. Less than one-half of the teachers (9 of 21) offered a favorable preliminary recommendation for further consideration of the possible adoption of *CR*. Most of these teachers supported their positive recommendations by emphasizing the content of the program (decoding and comprehension) and that the *CR* design addressed the needs of a wide range of students (e.g., grade levels, exceptionality). The six teachers not recommending further consideration of *CR* preferred to continue using their present remedial programs in their schools, expressing the concerns that *CR* was not aligned with the needs of their students and that *CR* was not supported by research. One additional teacher did not offer a recommendation because she felt no single program could meet the needs of all students. This teacher was counted as a negative recommendation, increasing the number of teachers not recommending *CR* from 6 to 7 (of 21). The five teachers that indicated they were uncertain reported that they needed additional time to investigate the program in greater depth before making a preliminary recommendation of any kind.

Teacher perspectives relating to the evaluation of CR. In making their recommendations, the nonprogram teacher judgments represented a combination of evaluation perspectives and program features. In turn, these perspectives provided a basis for suggesting how their recommendation decisions might relate to actual *CR* design features.

Table 7 summarizes teacher perspectives for the design of an ideal remedial reading program. As indicated in Table 7, of the three categories of features identified, teacher-centered were mentioned most frequently, with program-centered and assessment being less frequent. Most of the features suggested represented either constituents of instructional programs teachers were presently using or a generic set of features common to any basal reading program. None of the features reflected abstract principles underlying the design of a *DI* reading program (e.g., Carnine, Silbert, Kame'enui, & Tarver, 2003).

Table 8 contrasts the "best practices" for an ideal remedial reading program reported by teachers with the underlying design features of *CR*. As Table 8 shows, these teacher-identified features represented a substantially different perspective than those that underlie the design of *CR*.

Finally, Table 9 relates weaknesses of *CR* reported by teachers as evaluative criteria that influenced their recommendations for *CR*, but which represented misconceptions of actual *CR* design principles. As Table 9, shows, these weaknesses are logically consistent with the ideal features and best practices perspectives reported by teachers shown in Tables 7 and 8.

Summary and Discussion of the Results for Part 3

Despite the fact that 9 of 21 teachers (43%) made positive preliminary recommendations with regard to further considering the possible adoption of *CR*, 12 of 21 (57%) did not. In general, the overviews summarizing teacher perspectives on remedial reading program features and best practices shown in Tables 7, 8, and 9 are suggestive of a minimal understanding of *CR* design principles by all teacher participants, regardless of whether their *CR* recommendations were positive. Rather, the perspectives reported by teachers were consistent with the educational materials typically used by schools, which presumably represented teachers' prior experience. While understandable, this is a cause for concern with regard to (a) whether appropriate criteria were applied in evaluating *CR* and, if ultimately adopted, (b) what perspectives could be applied in support of the implementation of *CR* with the degree

of fidelity that would result in desired student achievement outcomes (e.g., those reported in Part 1). While the short 50-minute exploration used in this simulation was not sufficient for developing an in-depth understanding of *CR* design, the major concern raised by Part 3 of the study is that, within a school reform initiative, teachers involved in decision-making may hold prior-knowledge perspectives that preclude either their consideration of or support of effective adoption or implementation of *CR*. In fact, the perceived weaknesses of *CR* reported by teachers were among the major principles that make it effective (e.g., Engelmann & Carnine, 1982).

Discussion

Considered from the point of view of providing a foundation for advocating adoption and subsequent sustainability of *CR* within systemic school reform initiatives, the pattern of findings from Parts 1, 2, and 3 raise important considerations. Conducted in a rural, eastern North Carolina setting, Part 1 demonstrated the effectiveness of *CR* in accelerating the

Table 7

Features Identified by Teachers for Ideal Remedial Reading Programs

Features		
Teaching Strategies	Instructional Content	Assessment Orientation
1. Use *Reading Recovery* teaching techniques (3)	Identify words and word-chunks (1)	Determine the needs of students (4)
2. One-on-one teaching (1)	Daily writing, reading (oral and silent) (1)	Evaluate student progress (4)
3. Teacher an independent decision-maker focusing on the needs of the student (1)	Phonics (2)	
4. Small-group teaching (1)	Comprehension (2)	
5. Balance of teacher-centered and student-centered activities (2)	Wide variety of reading materials (trade books, newspapers, poems) (2)	
6. Peer tutors (1)	Colorful and attractive materials (1)	
7. *Corrective Reading* (1)		
8. Student allowed to progress at own rate (1)	Multiple instructional levels (1)	

Note. Numbers in parenthesis indicate how often this feature was mentioned.

reading achievement of previously low-performing (and low-SES) students. Given Part 1, Part 2 demonstrated through nonprogram teacher judgment that the instructional content of *CR*, even as limited to Levels A, B1, and B2 in this study, represented an advanced level of student performance that nonprogram teachers believed was not accomplished by a majority of their students receiving traditional instruction. In this regard, the nonprogram teachers judged the grade-level appropriateness of *CR* content as grade-level advanced, that is, as representing learning outcomes appropriate for students at higher grade levels and appropriate for high- and average-ability students rather than the low-ability students engaged in *CR* in grades 3 and 4 in Part 1. Together, the findings of Parts 1 and 2 are

Table 8

Teacher Perspectives for an Ideal Remedial Reading Program Contrasted with Corrective Reading (CR) *Design Principles*

Remedial Reading Program "Best Practices"	*Corrective Reading* Design Principles
1. Individualized teaching.	Because *CR* is effective with virtually all students who place, individualization—as the term is commonly used—is not relevant to effective instruction (rather, the concept of grouping for instruction is relevant for *CR*).
2. Teachers as independent decision-makers.	*CR* does require teachers to make very sophisticated decisions and judgments that are atypical—particularly in the areas of student error correction, student mastery, and student placement. However, teachers not highly experienced with *CR* and not proficient with the design model are discouraged from making curricular modifications.
3. Peer tutoring.	The *CR* philosophy is to design effective instruction that teaches higher-order concepts and skills directly. The only application for peer tutoring in this context would be to have the programs taught by peers rather than teachers, which makes little sense in most instances dealing with high-need students. Peer tutoring in itself is not considered an effective teaching strategy for most learning outcomes.
4. Students progressing at their own rate.	The *CR* design philosophy is to accelerate the rate of student learning progress. As long as the program is well designed and well taught, all students placed in the program make rapid educational progress.
5. Use of colorful and attractive materials.	The program purposefully eliminates color and other details when they are not relevant to the concepts or skills taught. In teaching decoding skills, every effort is made to focus student attention on the grapheme features necessary for mastery of decoding skills rather than force the students to "find" them embedded within irrelevant details.

Table 9

Teacher-identified Weaknesses of Corrective Reading *Representing Misinterpretations of* Corrective Reading *Design Principles*

Identified Program Weaknesses	Actual *Corrective Reading* Design Principles
1. Too much drill and practice.	*CR* includes the substantial amounts of practice that are required for mastery of any skill.
2. Progressively too difficult material.	*CR* minimizes the difficulty of new material by ensuring students gain proficiency in all prerequisite skills through the program prior to introducing new material.
3. Vocabulary is too varied.	CR systematically introduces and reviews all vocabulary taught in a variety of application contexts.
4. Too much emphasis on phonics-based instruction.	The *CR* design specifically teaches students decoding strategies as skills. Thinking and comprehension skills are addressed in other components of the program in concert with decoding skills as they are developed.
5. Students are passive learners.	*CR* is highly interactive and involves students as active learners throughout. The *CR* design uses the high rate of interaction to provide students with rapid positive feedback or error corrections.
6. Scripted lessons are not desirable.	*CR* does require teachers to follow scripted lessons. However, this ensures that the instruction students receive is as flawless as possible. Also, the "scripts" are designed to be relatively easy for teachers to use on a consistent basis as student learning evolves.
7. Instruction is too teacher-centered.	*CR* is teacher-centered in the sense that the teacher takes leadership in introducing, modeling, and guiding student learning of what is taught. This is considered by Direct Instruction (DI) design to be the most effective model for accelerating student learning across time.
8. Lacks individualization of instruction.	Because *CR* is highly effective for all students placed in the appropriate level, it is clearly individualized from the perspective of the learner. Thus, the idea that all students cannot learn from well-designed instruction is not accepted by the DI model that underlies *CR*.
9. Not effective for students with attention deficit disorder.	*CR* is validated as effective for all students placed appropriately in it. In fact, Part 1 of the present study included one instructional group composed of students with attention deficit hyperactivity disorder and emotional behavioral disabilities. In *CR*, there is no difference in effectiveness between regular students and those with "disabilities" as long as the students are placed in *CR* correctly.

mutually supportive of the potential systemic value of *CR* in school improvement initiatives.

In contrast, however, the findings in Part 3 represent perspectives that are contrary to the implications from Parts 1 and 2. More specifically, the findings in Part 3 raise serious concerns regarding how the potential value of *CR* can be communicated effectively to teachers and other school decision-makers and the means through which the benefits resulting from *CR* implementation can be advocated effectively. Despite positive recommendations for adoption by 9 of the 21 teachers in this study, none of the 21 teachers displayed any prior professional knowledge that served as a foundation for the evaluation of *CR*. Rather, the perspectives held by teachers in this study were contrary to the major design features that make *CR* effective. Of equal importance, these prior perspectives also raise concerns regarding the eventual implementation fidelity (and sustainability) of *CR* if adopted.

As Table 9 showed, nonprogram teachers not only substantially misinterpreted the design features of the program (negatively) but also advocated characteristics of "ideal" remedial reading programs that were contradictory to the research-validated *CR* program. Consistent with Hirsch's (1996) argument, teacher evaluation of the major design characteristics of *CR* reflected the "best practices" advocated by the established educational paradigm (see Kuhn, 1996). Therefore, considered together, the pattern of findings of this study from Parts 1, 2, and 3 suggests that teachers are able to respond appropriately to the specific student curricular content of *CR* (e.g., concrete examples of student instructional tasks) that otherwise might be ignored in discussions emphasizing the process of instruction in *CR*. As a result, in school-based reform contexts, without such concrete referential examples, teachers may reject *CR* as a promising reform alternative because of its lack of consistency with what Hirsch has labeled the accepted educational "thoughtworld," despite the substantial scientific and practical validity of *CR* (see Przychodzin-Havis et al., 2005).

Although this study has a number of limitations due to scope of sampling schools, teachers, and levels of *CR* content that should be addressed through replication and extension, the pattern of findings from Parts 1, 2, and 3 are suggestive of important priorities for both future research and practice. From the standpoint of research, these results call for a more detailed analytic documentation of the processes through which both remedial and developmental reading programs are adopted by schools. One key focus of such research should be on building an understanding of the semantic (i.e., ontological) foundation (see Hirsch, 1966; Sowa, 2000) of educational practitioners with the goal of developing approaches through which the potential systemic benefits of *CR* (and other DI programs) can be communicated effectively. Although replication and extension of the present study can contribute to such an initiative, such research is best framed within an advocacy framework for *CR* that explicitly relates to the sustainability and expansion issues (e.g., establishing institutional value) emerging from recent research on scale-up (see Romance & Vitale, 2007; Vitale & Romance, 2005).

In contrast to the research, implications of this study for practitioners using *CR* are straightforward and based on a rationale that follows directly from established *DI* principles (see Engelmann & Carnine, 1982). Consistent with Vitale and Joseph (in press), practitioners should work to communicate with other educators and parents in terms of the specific curricular tasks that provide concrete examples of the student learning outcomes resulting from being in *CR*. As an illustration, the design of Part 1 of this study, which used the preceding year's students as controls, allowed the effect of *CR* to be interpreted as something that could have benefited students if implemented a year earlier. Such examples provide a potential means of unambiguous communication in

terms of basic form sensory concepts (Engelmann & Carnine, 1982) that cannot be accomplished through presently accepted educational jargon (see Hirsch, 1996). Moreover, applying suggestions by Vitale and Joseph (in press), constructing displays consisting of representative examples of *CR* decoding and comprehension tasks across lesson sequences (and levels) would provide a powerful framework for advocating the educational benefits of *CR* in a fashion that also implies achievement comparisons with curricular outcomes of other instructional alternatives.

Although the preceding does recognize implicitly that achievement objectives established by state or local accountability systems must be accomplished through *CR* if it is to be considered effective by practitioners, it also is important to emphasize that such measures are not adequate for representing either the rich curricular outcomes engendered by *CR* or the potentially positive educational prognosis of remedial students who complete Level C of *CR-D* and *CR-C*. Although the extensive review of *CR* by Przychodzin-Havis et al. (2005) found consistent positive results for *CR* in more than 90% of 28 different studies, none of the *CR* studies reported achievement comparisons of *CR* students who completed Level C in both *CR-D* and *CR-C* with controls. In fact, the importance of studies in which students complete *CR-D* and *CR-C* implies a natural linkage of future *CR* research and practice. To address the major directions for future research in *CR* suggested by Przychodzin-Havis et al. (2005), who emphasized the need for broadening the demonstrations of the impact of *CR* on student performance on state accountability measures and in content subjects (e.g., social studies, science), the experimental designs with the greatest power will require student completion through Level C of *CR-D* and *CR-C* (cf. Torgesen et al., 2006). Keeping this perspective in mind while pursuing the priorities for research and practice suggested above has the potential to establish a continuing evolution through which the adoption of *CR* as a systemic solution to recognized educational problems can be advanced.

References

Biancarosa, G., & Snow, C. E. (2004). *Reading Next—A vision for action and research in middle and high school literacy: A report from the Carnegie Corporation of New York.* Washington, DC: Alliance for Excellent Education.

Carnine, D. W., Silbert, J., Kame'enui, E. J., & Tarver, S. (2003). *Direct instruction reading* (4th ed.). New York: Prentice Hall.

Carnevale, A. P. (2001). Help wanted...College required. Washington, DC: Educational Testing Service, Office for Public Leadership.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, New Jersey: Lawrence Erlbaum Associates Inc.

David, J. (1995-1996). The who, what, and why of site-based management. *Educational Leadership, 53,* 4-9.

Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications.* New York: Irvington Press.

Engelmann, S., Carnine, L., & Johnson, G. (1999). *Corrective Reading: Word attack basics, teacher's presentation book & guide: Decoding A.* Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., Haddox, P., Hanner, S., & Osborn, J. (1999a). *Corrective Reading: Comprehension skills, teacher presentation book: Comprehension A.* Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., Hanner, S., & Johnson, G. (1999). *Corrective Reading series guide.* Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., Meyer, L., Carnine, L., Becker, W., Eisele, J., & Johnson, G. (1999b). *Corrective Reading: Comprehension skills, teacher presentation book: Comprehension B1.* Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., Meyer, L., Carnine, L., Becker, W., Eisele, J., & Johnson, G. (1999c). *Corrective Reading: Comprehension skills, teacher presentation book: Comprehension B2.* Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., Meyer, L., Carnine, L., Becker, W., Eisele, J., & Johnson, G. (1999d). *Corrective Reading: Decoding strategies, teacher's presentation book and guide: Decoding B1.* Columbus, OH: SRA/McGraw-Hill.

Engelmann, S., Meyer, L., Carnine, L., Becker, W., Eisele, J., & Johnson, G. (1999e). *Corrective Reading: Decoding Strategies, Teacher's Presentation Book and*

Guide: Decoding B2. Columbus, OH: SRA/McGraw-Hill.

Hirsch, E., D. (1996). *The schools we need and why we don't have them.* New York: Bantum Doubleday Dell Publishing.

Kaniuka, T. S. (1997). Impact of improving student academic achievement on teachers' instructional expectations and decision-making: Implications for the school reform process (Doctoral dissertation, East Carolina University, 1997). *Dissertation Abstracts International, DAI-A, 58/02,* 352.

Kuhn, T. K. (1996). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive abilities test (CogAT).* Riverside, CA: Riverside Publishing.

National Assessment of Educational Progress (NAEP). (2002). *The nation's report card: Reading 2002.* Washington, DC: National Center for Education Statistics.

National Assessment of Educational Progress (NAEP). (2005). *The nation's report card: Reading 2005.* Washington, DC: National Center for Education Statistics.

National Center for Education Statistics. (2003). *Remedial education at degree-granting postsecondary institutions in fall 2000.* (NCES 2004-010). Washington, DC: U.S. Department of Education.

North Carolina Department of Public Instruction. (2004). North Carolina standard course of study. Retrieved January 4, 2009, from http://www.ncpublicschools.org/curriculum/languagearts/scos/2004/

North Carolina Division of Accountability Services. (2006). North Carolina end-of-grade reading comprehension tests grades 3-8. Retrieved January 4, 2009, from http://www.ncpublicschools.org/accountability/testing/eog/reading/

Przychodzin-Havis, A. M., Marchand-Martella, N. E., Martella, R. C., Miller, D. A., Warner, L., Leonard, B., et al. (2005). An analysis of *Corrective Reading* research. *Journal of Direct Instruction, 5,* 37-65.

Plisko, V. W. (2003). *The release of the National Assessment of Educational Progress (NAEP). The nation's report card: Reading and mathematics 2003.* National Center for Education Statistics. Washington, DC: U.S. Department of Education.

Romance, N. R., & Vitale, M. R. (2007, April). *Elements for bringing a research-validated intervention to scale: Implications for leadership in educational reform.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Snow, C. E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension.* Santa Monica, CA: RAND.

Sowa, J. F. (2000). *Knowledge representation: Logical, philosophical, and computational foundations.* New York: Brooks Cole.

Torgesen, J., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., Mansfield, W., et al. (2006). *National assessment of Title I: Interim report. Volume II: Closing the reading gap: First year findings from a randomized trial of four reading interventions for struggling readers.* Jessup, MD: U.S. Department of Education. (ERIC Document Reproduction Service No. ED491144)

Vitale, M. R., Boldt, J. F., Kaniuka, T. S., & Scott, B. J. (1999). Integrating research with school action: A collaborative strategy for paradigmatic change through systemic instructional improvement. In W. S. Thomson and C.L. Penny (Eds.), *Excellence through partnerships: Research in action* (pp. 160–187). Greenville, NC: East Carolina University.

Vitale, M. R., & Joseph, B. L. (in press). Broadening the institutional value of Direct Instruction implemented in a low-SES elementary school: Implications for scale-up and school reform. *Journal of Direct Instruction.*

Vitale, M. R., & Romance, N. R. (2005, April). *A multiphase model for scaling up a research-validated instructional intervention: Implications for leadership of systemic educational reform.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.