A Building-Based Case Study of Evidence-Based Literacy Practices:

Implementation, Reading Behavior, and Growth in Reading Fluency, K–4

Charles R. Greenwood, Yolanda Tapia, Mary Abbott, and **Cheryl Walton,**
Juniper Gardens Children's Project, University of Kansas

This study investigated the multiyear effects of a school-wide implementation of evidence-based literacy practices and a program to prevent early reading failure in one elementary school. Guided by a collaborative partnership/professional development model, the researchers hypothesized that teachers would implement and sustain their use of a range of new evidence-based practices and that these practices would produced accelerated levels and rates of growth in classroom reading behaviors broadly across students and in curriculum-based measurement (CBM) reading fluency. Results over 3 years indicated the following: (a) teachers did implement new evidence-based practices; (b) use of these practices with kindergarten and first-grade cohorts was associated with larger slopes in silent reading in second grade, a common point in time, compared to an older third-grade cohort not exposed to these strategies and to students at risk and with disabilities, who did not differ in their levels of classroom reading behavior; (c) classroom reading behaviors occurred most often in the presence of peer tutors, reading partners, or teacher-led one-on-one, small-group, or independent instructional arrangements as compared to entire group, teacher-led instruction; (d) growth in reading fluency was substantial overall; however, comparison of cohorts' progress at second grade indicated no differences in CBM fluency growth associated with students' differential histories of exposure to evidence-based practices, whereas differences in growth as a function of level of risk were found. Students at high risk progressed more slowly in attaining reading fluency than did typical students and low-risk students. Implications are discussed.

A gap exists between research and practice in special and general education literacy instruction (e.g., Greenwood & Abbott, 2001). For example, a number of recent reports have identified reading instruction practices supported by convincing evidence that they accelerate progress in learning to read (Burns, Griffin, & Snow, 1999; Snow, Burns, & Griffin, 1998); yet, the uptake of these practices as evidenced by use in local schools is much slower than one would wish (e.g., Lyon, 1999). Some of the often-cited solutions for reducing the gap have included (a) increasing the collaboration between researchers and teachers to make educational research more convincing, usable, and accessible to classroom teachers (Gerseten, Morvant, & Brengelman, 1995); (b) aligning practice with current research findings and accelerating the translation of research knowledge for practice (Carnine, 1997; Simmons, Kuykendall, King, Cornachione, & Kameenui, 2000); (c) creating professional roles in local schools for research lead teachers, professionals whose duties include identifying and translating research into practice through work with local teachers (Logan & Stein, 2001); (d) implementing professional development models that go beyond the traditional one-shot inservice teacher training to effect change in practice (Boudah & Knight, 1999; Boudah, Logan, & Greenwood, 2001; D. Fuchs & Fuchs, 2001; Vaughn, Hughes, Klingner, & Schumm, 1998). The questions as to whether such reforms will actually change teachers' practices, such changes will lead to improvements in students' reading processes and products, and implemented practices will be sustained over time in local schools still remain to be answered.

Serious challenges confront this demonstration. First, we know that changing teacher practices in local schools is far from easy (Boudah et al., 2001; Simmons et al., 2000), requiring years rather than months (Gersten et al., 1995). Reports have made it clear that changing teachers' practices through didactic classroom training is unlikely but that practice does improve with the addition of consultant monitoring of actual

Address: Charles R. Greenwood, 650 Minnesota Avenue–2nd Floor, Kansas City, KS 66101; e-mail: greenwood@ku.edu

classroom implementation accompanied by feedback to teachers (Boudah et al., 2001; Vaughn et al., 1998). Second, changing literacy instruction to an evidence-based approach is hampered by a lack of knowledge regarding exactly how to combine multiple effective practices into a comprehensive instructional program. For example, in a beginning reading program, evidence-based strategies such as *Reading Mastery,* Peer-Assisted Learning Strategies, and Sound Partners could be used separately or combined within a comprehensive instructional program based on a local school context and its guiding principles of effective reading instruction (Baker & Smith, 2001). Configuring the role to be played by several evidence-based practices within a comprehensive reading program over multiple grade levels is challenging because the breadth and depth of each practice must be factored into the overall equation. Third, measurement of progress in a reformed curriculum is too often lacking, or if available, insensitive. As a result, reforms in practices too often occur without evidence that (a) implemented practices meet standards of fidelity, (b) these practices are working, and (c) if current practices are not working, whether new changes are working (Deno, 1997; Greenwood & Maheady, 1997).

This article reports the outcomes of a multiyear case study of a local elementary school reform effort to promote teachers' use of evidence-based literacy practices (Abbott, Walton, Tapia, & Greenwood, 1999). The conceptual framework guiding this research posited that literacy skills are influenced by the everyday interactions each student has with the teacher and the curriculum during classroom instruction. Accelerated growth in literacy skills is seen to be a product of instruction defined by use of evidence-based practices that leads to short-term outcomes in reading aloud and silent reading behaviors, which, over time, lead to outcomes in terms of curriculum-based measurement (CBM) reading fluency. The relative effectiveness of classroom instruction overall is moderated by the fidelity of implementation of evidence-based practices and by professional development experiences that promote implementation and fidelity.

In this study, we investigated the multiyear effects of a professional development model, which was implemented in one elementary school and was designed to provide an ongoing process whereby evidence-based literacy practices could be routinely translated for practice in the context of actual classroom implementation. Although considerable evidence exists that CBM reading fluency is accelerated by various instructional interventions (e.g., L. S. Fuchs & Fuchs, 1986; L. S. Fuchs, Fuchs, Hamlett, & Bentz, 1994), observational data on students' early reading behavior during instruction and CBM progress in learning to read have rarely been reported (Chard & Kameenui, 2000). Our study further extended existing findings in the literature by examining multiyear—rather than single-year—effects on individual children's instructional experiences and growth in reading.

Based on prior research and the conceptual framework, the following hypotheses were put forth:

1. Teachers collaborating with researchers in a multiyear effort will implement and sustain the use of new evidence-based reading/literacy practices.
2. Differences in students' reading aloud and silent reading behaviors during instruction will be related to the use of these practices within grade cohorts over time, between cohorts with differential histories of exposure to evidence-based practices at second grade, by risk level, and for individual students with disability/minority language issues.
3. Instructional arrangements defined by one-on-one, independent, and small-group arrangements with peer tutors or the classroom teacher will promote more positive reading behaviors than will whole-class, teacher-led instruction focused on the entire group of students.
4. Differences will be observed in students' CBM reading fluency related to the use of these practices within grade cohorts' over time, between cohorts with differential exposure to evidence-based practices at second grade, by risk level, and for individual students with disability/minority language issues.

Method

Design

A longitudinal, sequential cohort design (Raudenbush & Chan, 1993) that incorporated process and product measures of growth in student performance was used. This design was appropriate because it is capable of incorporating repeated multiyear measurements, testing the effects of independent variables such as risk for academic delay, and using grade cohorts over time. Because of the research focus on young students who are just learning to read, only students who were in kindergarten, first grade, and second grade at the start of the project were included in the evaluative aspect. At the start of the project, three grade-level cohorts were identified for monitoring purposes over three school years: kindergarten (Cohort 1), first grade (Cohort 2), and second grade (Cohort 3). Cohort 1 was monitored for the period of K through second grade, Cohort 2 was monitored for the period of first grade through third grade, and Cohort 3 was monitored for the period of second grade through fourth grade. Because these cohorts shared second-grade experiences and measurements in common, it was possible in the sequential cohort design to compare the behavioral and CBM fluency progress made by each of the three cohorts at this overlapping point (Raudenbush & Chan, 1993). Questions about the accelerative effects of the early evidence-based interventions provided to Cohorts 1 and 2 were compared to Cohort 3, which served as a historical control group for this early set of experiences.

Participants

Students. This research took place in a small, urban elementary school with an annual student body that varied between 335 and 350 students in kindergarten through fifth grade. There were two classes per grade level per year. The students' racial makeup was as follows: 90% Euro-American/White, 7% Hispanic, and 3% African American/Black. In this Title I–qualified school, 41% of students received a free or reduced lunch. The special education model followed at this school was primarily inclusion in the general education classroom with limited use of pullout to a resource room in the school or in a nearby school. This school also was a professional development school in association with the school of education of a nearby university.

All students in each general education classroom participated in the program, but their actual instructional experiences varied, depending on their grade level and teachers' participation in each year of the study. For measurement tracking purposes in the design, 36 target students were identified in Year 1. They were stratified by three risk levels for early reading failure within each cohort. Based on their first-hand knowledge of the students, each kindergarten through second-grade classroom teacher selected two of his or her students as *lowest risk* ($n = 2$), *average risk* ($n = 2$), and *highest risk* ($n = 2$). This resulted in a total of 6 target students per classroom $\times$ 2 (same grade level classrooms per cohort) $\times$ 3 cohorts, for a total of 36 students.

The objective of this selection was to represent the full range of student diversity with respect to risk of early reading failure within each cohort (kindergarten through second grade). Students at lowest risk for early reading failure were children without disabilities who were high performing as indicated by the classroom teacher. Students at average risk were children without disabilities and whom the teachers judged to be average learners. Students at highest risk consisted of children with Individual Education Programs (IEPs) or with limited English proficiency (LEP), who were judged to be lowest achieving by the teacher.

These 36 students received the intervention program, as did all other students in the school. In addition, however, they received systematic repeated observation and CBM measurements over 3 school years unless they moved from the school ($n = 9$). In these cases, replacements at equivalent levels of risk were designated from among nontargeted students who had been involved in the program in all prior years. By the end of the study, 44 students (14–15 per cohort) had completed some or all of these assessments. Cohort 1 contained 1 girl (Student 38) with LEP and 1 boy (Student 39) with a visual impairment; Cohort 2 contained 1 boy (Student 18) with a behavior disorder. Cohort 3 contained no students with IEPs or LEP. A visual accommodation received by Student 39 was use of materials with large print.

Teachers. Twelve kindergarten through fifth-grade teachers and the principal participated annually, for a total of 16 teachers. This included 4 teachers who replaced teachers who left service at the building some time during the 3 years. An itinerant inclusion facilitator and a resource room teacher both worked part time at the school and participated each year. The median age range of the general education teachers was 40 to 44 years. Fifteen were women, and 1 was a man. Fifteen were European American/White, 1 was Latino. In terms of educational experience, 5 teachers had hours beyond the bachelor's degree, 3 held the master's degree, 7 had hours beyond the master's degree, 1 held an educational specialist degree, and 1 had completed a doctorate by the end of the study. Ten teachers (59%) had taught for more than 15 years, 2 had taught for 3 years or less, and the 5 remaining teachers' experience ranged from 4 years to 14 years.

Researchers/Consultants. A team of four to five researchers and staff at the Juniper Gardens Children's Project participated over all 3 years. The number of research staff present and available at the school to work with teachers was three, two, and one in Years 1 through 3, respectively. The team consisted of a career special education researcher, two postdoctoral associates, and two graduate students working on their doctorates. This team met on a weekly basis to plan and implement the professional development and measurement components of the project. Each staff member was assigned a caseload of teachers at the school with whom they collaborated and mentored during the project. Throughout the project, research staff were given space in the school, and at least one staff member was at the school 4 to 6 hours daily.

Partnership and Professional Development Procedures

A professional development model linking *partnership* and *collaboration* was used to guide activities within the school (Abbott et al., 1999). The model was designed to create a multiyear context that enabled researchers and teachers to work together in a sustained problem-solving process that would lead to evaluations of instructional problems, solutions, and replication of these solutions across teachers and classrooms.

Initial Planning. Planning occurred for approximately 4 months prior to implementing strategies at the school in Year 1. Following an overview presentation of the model to several school principals, an initial discussion concerning the project was held with an interested school principal. At the request of the principal, further discussions occurred with the building's site-based management council, which was composed of faculty member representatives. Faculty representatives reviewed these plans, polled their colleagues, and in agreement with the principal, decided in favor of participation in the project. After a period of collaboration had taken place at the school that demonstrated to both parties the existence, acceptance, and functionality of the partnership, a mission statement was drafted, as follows:

> Our mission is to formulate a partnership between teachers and researchers for the purpose of working together to improve the quality of learning by facilitating the rate of acquisition and mastery of academic and social skills for all students.

Although this clearly was a general mission statement at the outset, over time and as the partners worked together, the mission became operationalized in terms of specific professional development activities, collaboration on issues, and formative measurement activities, which are described in the remainder of this report. The focus in Year 1 was on implementing a measurement plan, building the partnership, and implementing the professional development model.

Implementation in Years 1 and 2. Measurement for progress monitoring and implementation of new evidence-based instruction practices began in earnest by November (winter) of the first year. Professional development experiences provided to the teachers were twofold: (a) in-service offerings, such as Class-Wide Peer Tutoring (CWPT) or Writer's Workshop, followed by in-classroom consultation or (b) individual problem solving following a teacher's interest or lead, for example, helping the kindergarten teachers develop and implement phonemic awareness instruction. To achieve full classroom implementation, researchers routinely modeled new instructional procedures, allowing teachers to observe implementation prior to trying the procedures on their own (Walton, 1998).

A full faculty in-service was held in the spring of Year 1, and individual meetings and classroom consultations were conducted throughout the year (Abbott et al., 1999). In subsequent years, it was possible to use entire-school in-service times before and during the school year, and the research staff and/or teachers from other buildings skilled in the use of a particular practice served as instructors. The most productive professional development times proved to be in the classroom, during planning time, or immediately before or after school.

As a matter of course, teachers were observed before and after implementing any new practice, and fidelity measures indicative of their use of the practice were collected. These measures served two purposes: as a basis for feedback and improvement of implementation and for documentation that the practice had been implemented as designed. It was typical for the quality of implementation to more than double—to 85% or greater—after initial implementation and subsequent consultation.

Evidence-Based Practices. The practices selected and implemented by teachers and researchers are listed in Table 1 by cohort and grade/year used within each cohort. For a practice to qualify as "evidence-based," evidence in the form of improved student learning from at least one empirical study was required. Preferred practices were ones supported by findings from experimental or quasi-experimental studies; however, this evidence was not always available. Cited in Table 1 for each practice is either a key study or synthesis of studies supporting efficacy. As can be seen in the table, the set of practices implemented within each cohort across 3 years varied, ranging across (a) 9 strategies, including Shared Book Experience through Writer's Workshop (Cohort 1); (b) 12 strategies, including Phonemic Awareness through Reciprocal Teaching (Cohort 2); (c) 6 strategies, including Partner Reading and Writer's Workshop through Reciprocal Teaching (Cohort 3). Some practices were implemented by more than one teacher per grade, used within and across cohorts, and experienced by students in more than a single year.

Teachers in each year implemented some variations in these practices by adding or dropping new ones within and/or across years, based on their own interests and decisions. The set of practices that influenced any single student's reading performance therefore was clearly not unitary and varied over time. For a majority of students, however, instructional experiences were most heavily concentrated in Partner Reading, Writer's Workshop, and CWPT.

In terms of the full scope of work in the building, it was initially agreed that the partnership effort would provide a priority of time and resources to teachers and students in the three youngest grade cohorts. The central focus of the collaboration was an effort to implement and sustain a high-intensity early literacy intervention experience for the original kindergarten through second-grade cohorts as they progressed through the next three school years.

At the end of the each school year, grade-level meetings at the school helped the researchers and teachers assess what had been successful and define and prioritize areas of change for next year (Abbott, Walton, & Greenwood, 2002; Abbott et al., 1999; James, Abbott, & Greenwood, 2001). The success or failure of various strategies was discussed, as were possible new strategies based on recent intervention research reports. Efforts were made prior to the start of subsequent school years to match in-service offerings to suggestions made by teachers and to recent evidence supportive of a strategy.

Maintenance of the Model. Yearly changes in teachers, researchers, and key players required an annual process devoted to maintaining and renewing the partnership between the researchers and the teachers. Issues included changes in personnel and roles, policies that supported or hindered the partnership, responses to prior results, and changes in the levels of resources—including fewer researchers available for consultation. In Year 2, researchers and teachers began investigating how to maintain the effort in the face of reductions in personnel and increasingly limited resources. Resource issues required exploration of the effects of a more economical and less intensive set of collaborative activities. Preparation for this phase began at the end of Year 2's debriefing with each individual teacher and before the in-services prior to Year 3. An initial effort was made to create a research lead teacher role at the school from among participating teachers (Logan

TABLE 1. Cumulative List of Implemented Reading-Related Interventions

Intervention	Cohort[a] 1	2	3	Supporting evidence
Shared Book Experience	K			Teale & Sulzby (1987)
Phonemic Awareness	K	1		O'Connor, Henkins, & Slocum (1995)
Repeated Reading	K	1		Samuels (1979)
Initial Reading Blending	1	1		Perfetti (1987)
Early Intervention Reading	1	1		Taylor, Short, Frye, & Shearer (1992)
Partner Reading	1, 2	1, 2, 3	2, 3	Mathes, Fuchs, Fuchs, Henley, & Sanders (1994)
Word Family Books	1	1		Juel & Roper-Schneider (1985)
Dolch Words	1	1		Leibert (1991)
Writer's Workshop	1, 2	1, 2, 3	2, 3, 4	Scardamalia & Breiter (1986)
Reading Class-Wide Peer Tutoring		2, 3	2, 3, 4	Greenwood, Delquadri, & Hall (1989)
Spelling Class-Wide Peer Tutoring		2, 3	2, 3, 4	Greenwood et al., (1989)
Partner Reading Questions		3	3, 4	Mathes, Fuchs, et al., (1994)
Reciprocal Teaching		3	3, 4	Rosenshine & Meister (1994)

[a]Usage according to grade: K = kindergarten; 1 = first grade; 2 = second grade; 3 = third grade; 4 = fourth grade.

& Stein, 2001), but this proved unfeasible without additional resources. Thus, at the beginning of Year 3, researchers worked with each teacher to solidify his or her instructional plans for the year in ways that would maintain existing evidence-based practices.

In the fall semester of Year 3, the number of research personnel was reduced to one person at the school. Individual researcher meetings with the teachers dropped to twice a month, because the researcher was at the school for only 2 hours every other week. The progress-monitoring observational and CBM measurements that had been conducted in the spring were continued. Although contacts between teachers and researchers were reduced, written feedback based on the continuing formative evaluation measures were delivered to teachers for use in instructional decision making. Overall, this resulted in most teachers continuing to implement practices learned in prior years. For example, first-grade teachers continued teaching Phonemic Awareness, and fourth-grade teachers continued using Writer's Workshop. Most teachers did not implement any new practices, however.

Measurement

As mentioned previously, researchers set up a measurement program in Year 1 to inform teachers of the effects of their instruction on the students' behavior and growth in reading fluency. Measures of strategy implementation, observations of student behavior, and CBM reading fluency probes were used as indicators of instructional processes and products. These measures were repeated within each year to support instructional decision making.

Process measures of the changes in teacher practices included (a) the number of new strategies implemented by teachers and (b) direct observations of classroom ecology, teacher, and student behavior during reading instruction. The product measure was a reading CBM. CBM reading fluency assessments began in November of 1996 and continued through February of 1999. Strategies implemented by individual teachers were recorded and documented whenever they took place. Observations began in January of the first year and continued through December of the 1998–1999 school year. Depending on the school schedule and available resources, observations and CBM measures were conducted as frequently as every month or at least every 3 months.

Students' Cumulative Exposure to Strategies. The number of evidence-based strategies implemented by teachers with the help of researchers at the school was monitored over time. For each *teacher* this was a simple count of the *number of strategies;* for each *student* the *cumulative number of strategies* they experienced each year was calculated.

Reading Instruction Observations. The *Code for Instructional Structure and Student Academic Response: Mainstream Version* (MS-CISSAR; Carta, Greenwood, Schulte, Arreaga-Mayer, & Terry, 1988; Kamps, Greenwood, & Leon-

ard, 1991) was used. MS-CISSAR is a multi-event classroom observation measure. Individual events were organized under three categories of *ecology, teacher,* and *student,* and each category was further separated into subcategories. Student events were organized into three subcategories: academic, task management, and inappropriate responses. *Academic responses* were defined as active responses to academic situations, commands, and instructions. *Task management responses* were defined as enabling of academic behaviors, that is, behaviors that positioned a student to make an academic response when given an opportunity to do so. *Inappropriate responses* were defined as those behaviors that interfered with the occurrence of academic responding and task management.

Similarly, subcategories were used to group classroom ecological and teacher behavior event classes. The subcategories of ecology were the *setting, activity, task, physical arrangement,* and *instructional structure*. These events described the physical location (setting), the subject matter (activity), the type of materials or media (tasks), the seating arrangement (physical arrangement), and the instructional arrangement (instructional structure). The teacher subcategories were *definition, behavior, approval,* and *focus*. These events described who the teacher was (teacher definition), what the teacher was doing (teacher behavior), the teacher's use of approval or disapproval (approval), and to whom the teacher's behavior was directed (the observed student, others, or all students).

Portable notebook computers running the *Ecobehavioral Assessment Systems Software* (EBASS; Greenwood, Carta, Kamps, & Delquadri, 1993) were used by observers to conduct MS-CISSAR observations. Recording of events was prompted by the software and paced using momentary time sampling. An observer was asked to record one of the event categories (e.g., ecology, teacher, or student) every 20 seconds over the total time of the observation. At the first time sample, the observer was asked to record ecological events by selecting from the list of ecology choices contained in the MS-CISSAR taxonomy. Twenty seconds later, the observer was asked to enter teacher events by selecting from those contained in the taxonomy. Twenty seconds following that, the observer entered student events from those contained in the taxonomy. This cycle of event recording was repeated once every minute of observation.

Observations of 30 to 60 minutes occurred during times that the general education teachers indicated were devoted to reading instruction. The actual mean length of the completed observations was 39.8 minutes (*SD* = 16.7 minutes, range = 13–67 minutes). Analysis of the data confirmed that 99% of the assessed time occurred in the general education classroom, wherein reading accounted for 68% of the time, language arts for 20%, spelling for 8%, and "other activities" for 40%.

Observer Training. Five observers collected data for the project in Year 1, four observers collected data in Year 2, and three observers did the collecting in Year 3. Observers were trained using the procedures, materials, and media described by Greenwood et al. (1993). An observer coordinator supervised this training, which included conducting agreement calibration checks, observing classrooms, and monitoring interobserver agreement over time. Training included (a) studying event definitions from a training manual, (b) using a computer-assisted tutorial containing response feedback for building accuracy to apply the definitions in the recording of classroom scenarios, (c) practicing actual classroom observation, and (d) conducting interobserver agreement checks with the coordinator. An observer was certified as trained following this sequence and after obtaining three consecutive checks at a total agreement score of 90%.

Interobserver Agreement. Interobserver agreement was routinely assessed by having two observers simultaneously observe the same target student. Observers sat close enough to each other to synchronize the beginning and ending of their observations but far enough apart so as not to influence each other's recording. Individual agreement checks lasted for the entire length of the observation. Agreement statistics were calculated using a computer program that compared the two observation files collected on each observer's computer. The software asked for the file name from each observer's disk and analyzed the records using interval-by-interval "point" agreement. An agreement was defined as both observers recording the same event at the same momentary time interval. Two agreement indices—percentage agreement and kappa—were computed. Percentage agreement was calculated as the number of agreements/total intervals times 100. Kappa was calculated using the formula provided by Hollenbeck (1978), and in contrast to percentage agreement, it controlled for the probability of chance agreement. The mean per percentage agreement was 97.4 (*SD* = 1.0, minimum = 95.4, maximum = 99.1). The mean kappa was .90 (*SD* = 0.05, minimum = 0.82, maximum = 0.98).

Reading CBM. Measures of reading fluency—words read correct per minute and words read incorrect per minute—were used to measure student progress in the reading curriculum. The reading materials used were obtained from the University of Oregon CBM Network. These materials were a compilation of reading passages taken from the Silver Burdett Ginn *Reading Series* (Pearson, 1987). Each student was pulled out of his or her classroom to a quiet room or area in the school to read to the research staff for approximately 5 minutes. Each student was allowed 1 minute to read each of three different pages of text at his or her instructional level. All students were given the same instructions on each occasion.

If the student did not know a word or hesitated for 3 seconds, the assessor said the word for the student and marked it as an error. If the student did not read a word correctly, omitted the word, substituted other words, or transposed words, these also were marked as errors. If the student self-corrected within 3 seconds or inserted words, these were not marked as

errors. Hyphenated words were counted as one word and numbers written as numerals were counted as words when read within the context of the passage. The assessor used a digital timer to time each student. Errors were marked on a separate probe copy of the text marked with the number of words contained in each line. Each student returned to his or her classroom after reading all three probes. The assessor calculated the number of words per minute for each text probe read. We used the procedures recommended by Shinn (1989) to select the middle best performance as the final reading rate. This work was supervised by a coordinator who trained assessors and checked the conduct of their work.

Reading CBMs were collected for the target students of each cohort over 3 years, with the exception of Cohort 1 (kindergarten), where reading CBM measurement began a year later, in first grade. Overall, 396 CBM measurements were taken. Across individual target students, including replacements, the number of measurements ranged from 4 (1% of all the data collected) to 14 (3.5% of the data collected). The number of CBM measurements was 93, 137, and 166 for Cohorts 1, 2, and 3, respectively.

Statistical Methods

Practices implemented by teachers and reading CBM data were recorded originally on paper forms and entered into SPSS files for analysis by research staff. MS-CISSAR data collected on notebook computers were exported directly to SPSS for analysis. Simple descriptive statistics and graphic displays were used to display the data on implemented practices. We used *Z*-score tests of the difference between conditional and unconditional probabilities to investigate differences between specific classroom instructional arrangements and their relationship to students' reading behavior (Allison & Liker, 1982; Castellan, 1979). For these tests, MS-CISSAR observations of all target students were pooled into single analyses.

CBM reading fluency and MS-CISSAR reading aloud and silent reading indicators were graphed for each cohort by consecutive month of schooling, ranging from 1 month to 45 months for students in kindergarten through fourth grade to enable direct comparisons of cohorts' progress over grades and time in school. Hierarchical linear modeling (HLM; Bryk, Raudenbush, Cheong, & Congdon, 2000; Raudenbush & Bryk, 2002) was used to address research questions related to the effects of cohort and risk level on growth parameters over time. Level 1 HLM analyses were used to compute intercept, linear slope, and acceleration growth values for individuals and groups (unconditional effects). Level 2 HLM analyses were used to model the effects of cohort and risk level on growth—the conditional effects (e.g., Rogosa & Willett, 1985). The decision to use a linear model versus a curvilinear model in a particular analysis was made by testing the increase in fit to the data resulting from either including or not including an acceleration parameter using the likelihood ratio test (Raudenbush & Bryk, 2002). With respect to the modeling of reading aloud and CBM reading fluency, fit was significantly improved by including a quadratic acceleration component for reading aloud, $\chi^2(4, N = 43) = 42.12$, $p = .0001$, and CBM reading fluency, $\chi^2(4, N = 43) = 27.67$, $p = .0001$.

HLM is tolerant of missing data, and a unique advantage of HLM analysis is the ability to compute the mean level (i.e., intercept) at a single point in time and test for mean differences between groups at this point in time (Raudenbush & Bryk, 2002). Unless otherwise indicated, tests of the intercept means in this study were centered in the middle of the second-grade overlap shared by all three cohorts. It thus was possible to compare cohorts' relative progress at this common point in time. Analyses of the main effects of cohort (3) and risk level (3), but not of their interaction, were conducted due to the problem of relatively low cell sizes (four to five per cell).

Results

Did Teachers Implement New Evidence-Based Practices?

This hypothesis was accepted. Over the entire project, teachers implemented a total 13 different evidence-based strategies in collaboration with the researchers (see Table 1). For example, Partner Reading, Writer's Workshop, CWPT, and Reciprocal Teacher were used within and across cohorts in multiple years, showing replication and sustainability during the project. The mean number of strategies experienced by individual students in each year of the project was 3, 3, and 1 per student in Years 1, 2, and 3, respectively. The cumulative mean number of strategies actually experienced by students over 3 years was 7 per student ($SD = 1.5$). The cumulative mean number of strategies received per student per cohort was 8, 6, and 6 for Cohorts 1, 2, and 3, respectively. Students in Cohort 1 received an additional 2 strategies over the life of the project.

Observed Trends and Changes in Students' Reading Behavior

Results generally showed reading aloud emerging in kindergarten and first grade, declining thereafter as increases in silent reading in Grades 2, 3, and 4 were observed (see Figure 1). Also declining in later grades was the total time that students were observed (reading aloud + silent reading). The hypotheses that (a) the younger cohorts—1 and 2—would exceed Cohort 3 at second grade and (b) no differences would exist at second grade between risk groups due to differential histories of exposure to evidenced-based practices were accepted for silent reading only.

Cohort Effects. Growth parameters for reading aloud in the entire sample (unconditional effect) were 13.8%, −0.05%,

and –0.04% in percentage occurrence units for mean level at second grade (intercept), slope, and acceleration, respectively. These overall data indicated that reading aloud was declining at a linear rate of -0.05% occurrence per month in school plus an additional -0.04% per month. HLM Level 2 tests indicated significant cohort differences in reading aloud mean levels (intercepts), linear slope, and acceleration growth parameters (see Table 2). Cohort mean levels at second grade were 7.3%, 11.4%, and 15.6% for Cohorts 1, 2, and 3, respectively. All cohorts' slope and acceleration parameters were negative, replicating the overall effect in that the occurrence of reading aloud was generally declining over time (see Table 2). As illustrated in Figure 1 (see upper panel), reading aloud emerged in the last months of kindergarten, accelerated to the 19% level for most of first grade, and then dropped below 5% in second grade. Both Cohorts 2 and 3 produced reading aloud within the 12% to 23% range in their first year, but this gradually declined in subsequent years.

The entire sample mean level for silent reading at second grade was 9.6% occurrence, with a linear slope of 0.37% per month. Rather than declining like reading aloud, silent reading increased at a rate of just over one-third percentage occurrence point per month. Because the fit of the growth model was not improved by adding an acceleration component, only linear slope and intercept were needed to adequately describe growth in silent reading. Although no significant differences in cohorts' silent reading mean levels occurred at second grade, as hypothesized, the slope effect was significantly different (see Table 2). Growth in silent reading was most rapid for Cohort 1, at 0.67% per month compared to 0.47% and 0.27% for Cohorts 2 and 3, respectively (see Table 2). Silent reading emerged in first grade to within the 11% to 14% range but continued slightly lower into second grade (see Figure 1). Similar levels were observed in Cohorts 2 and 3 in their first year, and they continued to be relatively level into subsequent years. Of interest was the observation that even for the older cohorts, silent reading did not occur more than 16% of the total reading instruction time. Even though time spent in silent reading was increasing, total reading behavior (aloud plus silent) was declining (see Figure 1).

Risk Effects and Disability/ELL Effects. HLM analyses revealed no significant differences in growth parameters for either reading aloud or silent reading by level of risk (see Table 2). All three students in the disability/LEP category were engaged in reading behaviors during instruction that were at or above the levels for their cohort (see Table 3). Student 38, an English language learner, was reading aloud at levels above her cohort average and displayed silent reading during instruction at levels comparable to her cohort. Student 39, who had a visual impairment, was most prominently engaged in silent reading (see Table 3). Student 18, who had a behavior disorder, also displayed growth parameters substantially above those of his cohort in both reading aloud and silent reading.

One-on-One, Independent, and Small-Group Arrangements Versus Teacher-Led Group Instruction

In general, the hypothesis that reading aloud would occur most readily in the presence of one-on-one, independent, and small-group instructional arrangements with peer tutors or the classroom teacher was supported. The largest conditional probabilities were observed during reading activities when the student was using a reader in a one-on-one situation with the teacher (Cohort 1) or a peer tutor (Cohorts 2 and 3), yielding reading aloud probabilities of 0.53, 0.57, and 0.61, respectively (see Table 4). The unconditional or base-level probabilities of reading aloud in all other instructional conditions were comparatively lower at 0.11, 0.12, and 0.10, in order by cohort. The next best instructional arrangements for promoting reading aloud were small groups and the classroom teacher. Only occasionally was reading aloud promoted by arrangements that included worksheets, other media tasks, or whole-class grouping.

With respect to silent reading (see Table 5), the hypothesis also was supported, but whole-class, teacher-led instruction also was an active promoter in each cohort. The highest probabilities of silent reading occurred during reading instruction using (a) workbooks and whole-class instruction led by the regular teacher, $p = 0.32$ (Cohort 1), (b) workbooks and whole-class instruction with the regular teacher, $p = 0.26$ (Cohort 2), and readers and whole-class instruction led by a student teacher, $p = 0.72$ (Cohort 3). The unconditional (base level) probabilities of silent reading were comparatively lower at 0.08 (Cohort 1), 0.10 (Cohort 2), and 0.14 (Cohort 3). The other significant promoters of silent reading were small group, independent, or one-on-one with readers, workbooks, and worksheet tasks/materials. Unlike the promoters of reading aloud, who wanted it to be used only during reading instruction, silent reading was promoted by similar arrangements in language arts (Cohorts 1 and 3) and spelling instruction (Cohort 1; see Table 2).

Observed Trends and Changes in Growth in CBM Reading Fluency

In general, results indicated substantial growth in CBM reading fluency. HLM produced entire sample growth parameters of 58.3 words per minute (mean level at second grade), 3.1 words per minute per month (slope), and -0.05 words per minute per month (acceleration). Overall, fluency was growing, but with negative acceleration, over 3 years (see Table 2). The hypotheses that (a) the younger cohorts (1 and 2) would exceed Cohort 3 at second grade and (b) no differences would be found between risk groups due to differential histories of exposure to evidence-based practices for CBM reading fluency were not supported.

TABLE 2. Growth Curve Parameters for the Entire Sample, Cohorts, and Risk Levels

	Intercept at 2nd grade				Slope				Acceleration			
Effect	*M*	*t*	*df*	*p*	*M*	*t*	*df*	*p*	*M*	*t*	*df*	*p*
Reading Aloud												
Unconditional	13.829	9.41	42	0.000	–0.053	–1.08	42	0.285	–0.039	–5.34	42	0.000
Cohort Adj.	4.152	2.12	41	0.040	1.091	2.89	41	0.007	0.041	3.05	41	0.005
1	7.267				–1.959				–0.141			
2	11.419				–0.869				–0.100			
3	15.570				0.222				–0.059			
Risk Adj.	–1.352	–0.72	41	0.475	0.041	0.63	41	0.535	0.004	0.44	41	0.661
High	12.324				–0.010				–0.034			
Average	13.676				–0.051				–0.038			
Low	15.028				–0.091				–0.042			
Silent Reading												
Unconditional	9.649	16.54	42	0.000	0.367	6.77	42	0.000				
Cohort Adj.	–1.687	–1.57	41	0.124	–0.205	–3.34	41	0.002				
1	12.636				0.674							
2	10.949				0.469							
3	9.261				0.265							
Risk Adj.	0.612	0.89	41	0.381	–0.015	–0.25	41	0.804				
High	10.304				0.355							
Average	9.692				0.370							
Low	9.081				0.384							
CBM Read Fluency												
Unconditional	58.267	14.50	42	0.000	3.101	15.75	42	0.000	–0.045	–3.46	42	0.002
Cohort Adj.	6.516	1.34	41	0.189	0.117	0.32	41	0.752	0.006	0.25	41	0.801
1	51.249				2.738				–0.049			
2	57.764				2.855				–0.043			
3	64.280				2.972				–0.037			
Risk Adj.	21.065	5.28	41	0.000	0.483	1.99	41	0.053	–0.043	–3.06	41	0.004
High	40.132				2.715				–0.014			
Average	61.197				3.198				–0.056			
Low	82.261				3.681				–0.099			

Note. Adj. = Increment or decrement to the hierarchical linear modeling Level 1 coefficient reflecting the effect of cohort or risk on growth; CBM = curriculum-based measurement.

Cohort Effects. HLM analyses revealed no significant cohort differences in growth parameters (see Table 2), suggesting that relative to the mean level of Cohort 3, the early interventions experienced by the younger cohorts, 1 and 2, had not differentially accelerated their reading fluency levels by second grade. Visual inspection of the CBM data indicated that all cohorts were increasing their fluency levels over time (see Figure 2). In September of each year, performance started lower in most cases than in the prior year, but it increased over time with successive months in a school year.

Cohort mean intercept values at the last measurement occasion of the study (February 1999) were 58.8, 72.2, and 113.7 correct words per minute for Cohorts 1, 2, and 3, respectively. Projected final year-end values (May 1999) were 69.9, 80.9, and 120.9 correct words per minute.

Risk Effects. HLM Level 2 analyses revealed significant differences in CBM reading fluency growth parameters by level of risk (see Table 2). In general, high-risk students had the lowest mean level of reading fluency at second grade,

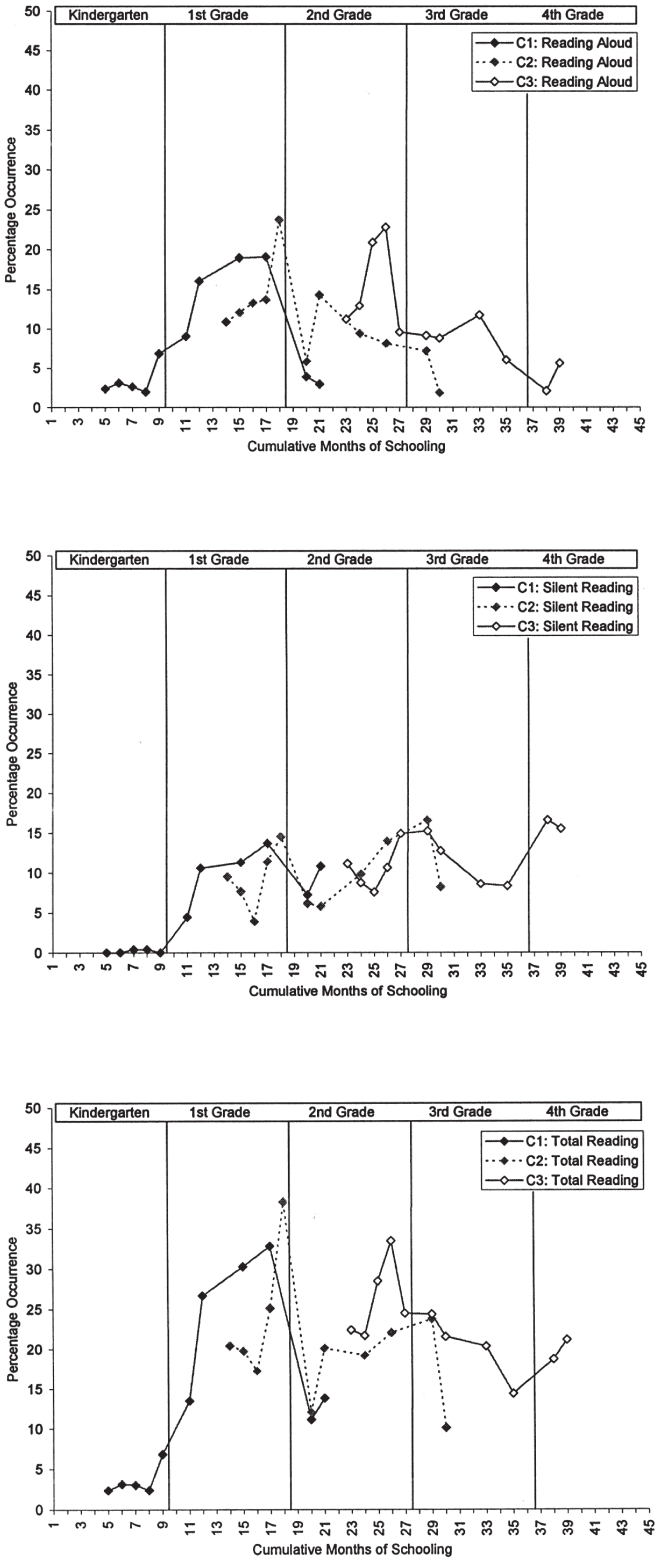

FIGURE 1. Trends in MS-CISSAR reading aloud (upper panel) and silent reading behaviors (middle panel) and in total reading (lower panel) by cohorts over consecutive months of schooling.

and they were making the lowest monthly progress as compared to average- and low-risk group students. The best news was the fact that the negative acceleration component for the high-risk students was smallest at –0.14 words, compared to –0.56 and –0.10 words for the average and low-risk students. Thus, the shape of their trajectory over time was more linear and accelerating as compared to those of the other two groups.

Disability/LEP Effects. Student 39, who had a visual impairment, was making CBM reading fluency progress comparable to that of other students in Cohort 1 (see Table 3). Students 38 and 18, however, were clearly struggling to improve their CBM reading fluency. Student 38, an English language learner, had a little more than half the mean level compared to Cohort 1 and a slope of less than one new word per month. Like Student 38, Student 18 had only half the mean level of his cohort, but a more promising slope of 1.7 words per month, better than previously but still below his cohort's slope of 2.9.

Discussion

In this case study, the processes and products of a multiyear elementary school reform effort were examined in terms of teachers' implementation of evidence-based practices, students' engagements in reading aloud and silent reading during instruction, and CBM reading fluency in kindergarten through fourth grade. At the outset, it was hypothesized that the building-based collaborative partnership between building faculty and university researchers would lead to multiple positive outcomes for practice, reading behavior, and reading fluency over the 3 years of application. The results generally supported this position. For example, the 3-year linear rate of growth in CBM reading fluency for the entire sample working in instruction-level material was 3.1 new words per month of schooling, ranging from a high of 3.7 for the low-risk students to a low of 2.7 for the high-risk students. HLM analyses indicated that these rates of progress were slightly decelerating over years and grade levels.

Most striking was the covariation between and among the use of new practices; the early onset acceleration in reading aloud and its subsequent deceleration, followed by sustained growth in silent reading; and growth in CBM reading fluency. These findings provide additional empirical support for a performance-based instructional framework in which implementation of evidence-based instructional practices designed to promote reading behavior during reading instruction is linked to students' growth in important academic outcomes such as CBM reading fluency. Additional support for the relationship between use of evidence-based practices and classroom reading behavior followed from the demonstration that probabilities of reading behavior were accelerated in the presence of one-on-one, small-group, and independent reading

TABLE 3. Growth Curve Parameters for Students with Disabilities

Unit	Reading aloud			Silent reading			CBM reading fluency		
	Intercept	Slope	Acceleration	Intercept	Slope	Acceleration	Intercept	Slope	Acceleration
Cohort 1	*7.267*	*–1.959*	*–0.141*	*12.636*	*0.674*		*51.249*	*2.738*	*–0.049*
Student 38–ELL	16.745	–6.344	–0.584	16.092	0.538		27.092	0.081	–0.145
Student 39–OHI	–1.716	–5.482	–0.428	31.079	2.296		48.285	3.385	–0.013
Cohort 2	*11.419*	*–0.869*	*–0.100*	*10.949*	*0.469*		*57.764*	*2.855*	*–0.043*
Student 18–BD	15.273	7.453	0.815	9.075	0.985		23.654	1.700	0.006

Note. ELL = English language learner; OHI = other health impairment; BD = behavior disorder; italics = cohort growth parameters; acceleration parameter for silent reading was not needed in the growth model.

arrangements involving peers. Consequently, practices that include these instructional arrangements should be used more frequently by teachers to create greater opportunities for reading behavior production in the classroom (Chard & Kameenui, 2000).

The hypothesis that teachers would implement a range of new evidence-based practices was accepted. Within cohorts and across time, variation in practices was seen to move from establishing emerging literacy skills, such as book and print concepts and phonemic awareness, in Cohort 1 to establishing story reading, spelling, and writing using practices such as Partner Reading, CWPT, Writer's Workshop, and Reciprocal Teaching in Cohort 3. By the end of the project, the students in Cohort 1 had experienced two more evidence-based strategies than had their counterparts in the other two cohorts. Overall, implementation was stronger and more intensive in the first 2 years of the project; it was less so in the third year because teachers typically continued using previously learned strategies rather than new procedures due to (a) a reduction in the researchers' time and effort at the school and (b) the lack of a local building facilitator in a position to promote new strategies that year.

HLM analyses of the entire sample at second grade indicated significantly greater than zero mean levels, slopes, and/or acceleration parameters for reading aloud, silent reading, and CBM reading fluency. For reading aloud, both slopes and acceleration parameters were significant and negative, indicating a rapidly declining trend over time. For silent reading, linear slope was positive and an acceleration parameter was not needed in the model. For CBM reading fluency, the slope was positive but acceleration was negative, indicating slowing progress over time. This finding was consistent with other reports indicating that reading aloud CBM growth rates tend to decline at higher grade levels (L. S. Fuchs, Fuchs, & Hamlett, 1993).

The hypothesis that growth in the classroom reading behavior for Cohorts 1 and 2 would exceed that of Cohort 3 by the middle of second grade due to the former's longer histories with and exposure to evidence-based practices was rejected for reading aloud but supported for silent reading. In the case of reading aloud, Cohort 3 had the highest level, with the most level trend over time, whereas Cohorts 1 and 2 had lower levels and decelerating trends. This outcome was just the opposite of what had been hypothesized.

In the case of silent reading, the cohorts' mean levels at second grade were not statistically different but were in the hypothesized direction; rates of growth were positive and significantly different in the hypothesized direction. Students in the younger cohorts appeared to be more rapidly replacing early growth in reading aloud with growth in silent reading, more so than in Cohort 3, where oral reading was continuing. This result was most likely due to the extensive use in Cohort 3 of CWPT, which has a major reading aloud component. In terms of total reading behavior (reading aloud + silent reading) during instruction, all cohorts were declining somewhat in second grade and thereafter (see Figure 1).

The hypothesis that instructional practices defined by one-on-one, independent, and small-group arrangements with peer tutors would promote more reading behavior compared to whole-class, teacher-led instruction that focused on the entire group of students was supported (Elbaum, Vaughn, Hughes, Moody, & Schumm, 2000; Logan, Bakeman, & Keefe, 1997). It was also clear that some similar and some different instructional conditions were effective in promoting silent reading. Compared to reading aloud, silent reading was more likely to occur in association with the use of task materials such as workbooks and worksheets as well as in whole-class, small-group, and one-on-one instructional groupings. Unlike the reading behavior and CBM analyses that were broken out by cohorts, risk groups, and disabilities, these conditional probability analyses were cumulative in that they focused on the entire group of students over all 3 years in order to represent the effects on overall reading behavior of all practices, teachers, classrooms, and years.

Although substantial growth was evident for Cohorts 1 and 2, the hypothesis that they would exceed that of Cohort 3 by the middle of second grade was rejected because growth parameters (level, slope, acceleration) for the cohort effect were

TABLE 4. Instructional Arrangements Promoting Oral Reading

Ecological model						Behavior			
Activity	Task	Instructional grouping	Teacher definition	Frequency	%	Frequency	Probability	*Z*	*p*
			Cohort 1						
Reading	Readers	One on one	Regular	76	1.8	40	0.53	10.979	0.001
Reading	Readers	Small group	Regular	546	13.1	162	0.30	12.586	0.001
Reading	Worksheet	Small group	Regular	193	4.6	41	0.21	4.302	0.001
Reading	Readers	Whole class	Regular	819	19.7	120	0.15	3.045	0.01
Uncondit. prob.				4158	77.2	450	0.11		
Total seq. rec.				5,386					
			Cohort 2						
Reading	Readers	One on one	Peer tutor	114	4.5	65	0.57	13.842	0.001
Reading	Readers	One on one	Regular	142	5.6	43	0.30	6.302	0.001
Reading	Readers	Small group	Regular	216	8.5	49	0.23	4.534	0.001
Reading	Other media	Small group	Regular	148	5.8	26	0.18	2.031	0.05
Uncondit. prob.				2531	73.9	296	0.12		
Total seq. rec.				3,423					
			Cohort 3						
Reading	Readers	One on one	Peer tutor	269	7.5	163	0.61	24.904	0.001
Reading	Readers	Independent	Regular	101	2.8	23	0.23	3.884	0.001
Reading	Readers	One on one	Regular	301	8.3	61	0.20	5.225	0.001
Reading	Readers	Whole class	Regular	384	10.7	59	0.15	2.986	0.01
Uncondit. prob.				3605	71.9	369	0.10		
Total seq. rec.				5,012					

Note. Uncondit. prob. = unconditional probability; Total seq. rec. = total sequence recorded.

TABLE 5. Instructional Arrangements Promoting Silent Reading

Ecological model						Behavior			
Activity	Task	Instructional grouping	Teacher definition	Frequency	%	Frequency	Probability	*Z*	*p*
			Cohort 1						
Reading	Workbooks	Whole class	Regular	57	1.4	18	0.32	6.199	0.001
Language	Other media	Small group	Regular	62	1.5	14	0.23	3.987	0.001
Reading	Readers	One on one	Regular	76	1.8	16	0.21	3.942	0.001
Reading	Readers	Independent	Regular	77	1.9	16	0.21	3.884	0.001
Reading	Readers	Small group	Regular	546	13.1	93	0.17	6.918	0.001
Reading	Worksheet	Small group	Regular	193	4.6	29	0.15	3.318	0.001
Spelling	Worksheet	One on one	Regular	207	5.0	25	0.12	1.974	0.05
Uncondit. prob.				4158	77.20	336	0.08		
Total seq. rec.				5,386					
			Cohort 2						
Reading	Workbooks	Whole class	Regular	91	3.6	24	0.26	5.007	0.001
Reading	Workbooks	Independent	Regular	35	1.4	9	0.26	3.014	0.01
Reading	Worksheet	Small group	Regular	102	4.0	22	0.22	3.763	0.001
Reading	Readers	Small group	Regular	216	8.5	33	0.15	2.515	0.05
Uncondit. prob.				2531	73.9	2246	0.10		
Total seq. rec.				3,423					
			Cohort 3						
Reading	Readers	Whole class	Student-Teacher	54	1.5	39	0.72	11.634	0.001
Reading	Readers	Independent	Regular	101	2.8	35	0.35	5.689	0.001
Reading	Readers	Small group	Regular	103	2.9	24	0.23	2.656	0.01
Reading	Readers	Whole class	Regular	384	10.7	85	0.22	4.354	0.001
Language	Workbooks	Independent	Regular	83	2.3	18	0.22	1.995	0.05
Reading	Readers	One on one	Regular	301	8.3	65	0.22	3.65	0.001
Uncondit. prob.				3605	71.9	488	0.14		
Total seq. rec.				5,012					

Note. Uncondit. prob. = unconditional probability; Total seq. rec. = total sequences recorded.

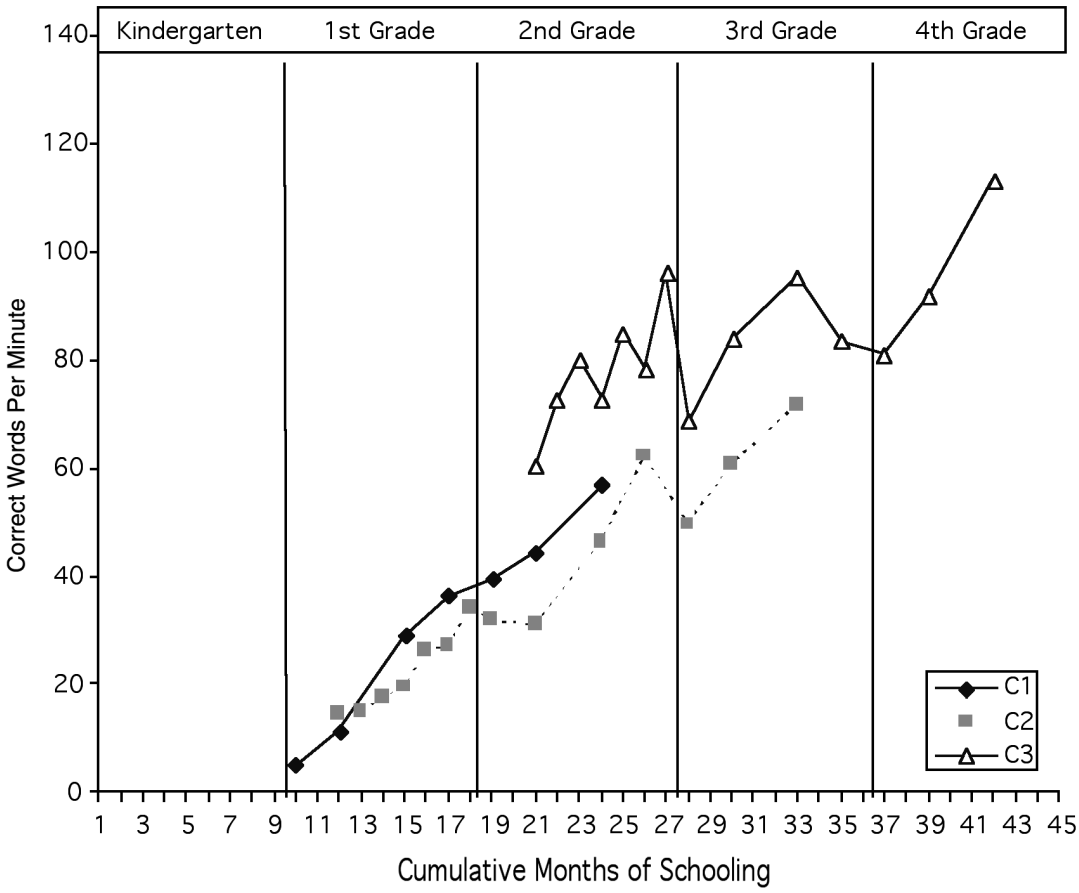

FIGURE 2. Trends in reading CBM correct words per minute by cohorts over consecutive months of schooling.

statistically equivalent (see Table 2). This was somewhat surprising, because evidence-based instructional factors historically differentiated the early experiences of each cohort. Cohorts 1 and 2, for example, had received phonemic strategies—including rapid letter naming, onset recognition, and segmentation—during their first year (Abbott et al., 2002), whereas Cohort 3 had not.

The hypothesis that students at greatest risk would evidence progress equivalent to that of typical and low-risk students was accepted for reading aloud and silent reading but rejected for CBM reading fluency. This was good news for students' engagement in classroom reading behavior, as prior research had shown high-risk students to be less engaged in academic responding, including reading aloud and silent reading, than low-risk students (Greenwood, Delquadri, & Hall, 1989; Greenwood, Hart, Walker, & Risley, 1994). These findings supported the conclusion that the evidence-based interventions were advancing reading behavior during instruction at comparable levels across risk groups, including students with disabilities. There was some good news for CBM reading fluency progress because the shape of the low-risk groups' growth trajectory was least slowing over time in comparison to the other two cohorts (see Table 2), even though the low-risk group's mean level and linear slopes were lower relative to those of the other groups.

Although considerable progress was made overall, these results indicated that the multiyear use of evidence-based practices was not yet sufficiently powerful to advance (a) the CBM reading progress of the two younger cohorts above that of Cohort 3 by second grade or (b) the progress of high-risk students to within that of typical and low-risk student groups by second grade. This was true even though it appeared that the interventions were advancing students' use of classroom reading (aloud and silent) in ways that had been hypothesized. We previously argued that accelerative effects in classroom processes and products such as these can only come from the use of instructional interventions where students grow faster in less time in school (Greenwood et al., 1994), and as noted in this case study, this goal clearly is not easily achieved. The development and implementation of instructional practices capable of producing these accelerative effects needs to be studied.

Limitations

An obvious limitation of this study was the lack of an experimental or quasi-experimental analysis of these effects. The

sequential cohort design, unlike these designs, does not control for a rival hypothesis, such as differential histories of instructional experiences, as was the case in this study, nor does it separate out growth due to maturation versus differences in instructional experiences. This design did provide important advantages over other potential case study designs:

- the inclusion of all students from different grade levels,
- a span of development over more than a single school year,
- the tracking of intervention differences and variations between cohorts and within cohorts over time,
- description of changes in processes and products of interest, and
- comparisons of growth between grade cohorts and risk levels at a common point in time and over time.

Implications and Future Research

The current findings support the effectiveness of professional development approaches that extend beyond in-service work to include sustained classroom consultation to effect changes in classroom practice. Also supported were features of collaboration focused on the interests and concerns of classroom teachers as related to their continuing participation in, planning of, implementation of, and evaluation of new practices. The current longitudinal findings demonstrated cumulative effects of literacy instruction over 3 years of school, clearly describing how teachers' annual variation in use of instructional practices affected individual students' learning histories with respect to reading. Students with a disability or English learning issues benefited equally well in terms of inclusion in the general education classroom reading program.

Although most teachers continued using practices in Year 3 that had been implemented in prior years, they did not increase the use of new strategies in the face of a reduced presence by researchers in that year. In addition, efforts to create a research lead teacher role in the building from among the existing school faculty were not successful. The importance of building this component in future efforts thus is noted here, along with related implications. First, too much of this project was school-based rather than district policy–based, resulting in partnership reforms that lasted only as long as did the researchers' external funding. For example, the researchers provided the formative measurement resources and personnel power in this project because they fell outside of the interest and resources of the school. Second, how the model would be continued at the school or the university was not an initial policy issue, although it should have been. Consequently, this issue was only addressed during the last year, when it appeared to be too late.

Future work on this and similar models must address these problems by making it possible and feasible for districts and schools to collect and analyze data and make instructional decisions using their own CBM data tools and resources (Baker & Smith, 2001; Good & Kaminski, 2000). Future work must also make it possible to sustain the collaboration between researchers and teachers (Boudah et al., 2001).

AUTHORS' NOTES

1. This work was supported by a directed research grant (No. H023G50012) from the Office of Special Education Programs (OSEP), U.S. Department of Education. Additional support was provided by Grant No. H029D60040, a leadership personnel preparation project. Opinions expressed herein are those of the authors and do not represent the position of the U.S. Department of Education.
2. The authors would like to acknowledge the significant contributions to this work of the principal, faculty, and students of the Turner Grade School, Turner District, Kansas City, Kansas. Additional thanks are owed Evelia Acosta, Harriett Dawson, Joe Delquadri, Shalini Reddy, Christine Rogers, and Rebecca Perez.

REFERENCES

Abbott, M., Walton, C., & Greenwood, C. R. (2002). Research to practice: Phonemic awareness implementation. *Teaching Exceptional Children, 34*(4), 20–26.

Abbott, M., Walton, C., Tapia, Y., & Greenwood, C. R. (1999). Research to practice: A blueprint for closing the gap in local schools. *Exceptional Children, 65,* 339–352.

Allison, P. D., & Liker, J. K. (1982). Analyzing sequential categorical data on dyadic interaction: A comment on Gottman. *Psychological Bulletin, 91,* 393–403.

Baker, S., & Smith, S. (2001). Linking school assessments to research-based practices in beginning reading: Improving programs and outcomes for students with and without disabilities. *Teacher Education and Special Education, 24,* 315–332.

Boudah, D. J., & Knight, S. L. (1999). Creating learning communities of research and practice: Participatory research and development. In D. M. Byrd & D. J. McIntyre (Eds.), *Research on professional development schools: Teacher education yearbook VII* (pp. 97–114). Thousand Oaks, CA: Corwin Press.

Boudah, D. J., Logan, K. R., & Greenwood, C. R. (2001). The research to practice projects: Lessons learned about changing teacher practice. *Teacher Education and Special Education, 24,* 290–303.

Bryk, A. S., Raudenbush, S. W., Cheong, Y. K., & Congdon, R. (2000). *HLM5: Hierarchical linear and nonlinear modeling.* Chicago: Scientific Software International.

Burns, M. S., Griffin, P., & Snow, C. E. (1999). *Starting out right: A guide to promoting children's reading success.* Washington, DC: National Research Council, Commission on the Prevention of Reading Difficulties in Young Children.

Carnine, D. (1997). Bridging the research-to-practice gap. *Exceptional Children, 63,* 513–521.

Carta, J. J., Greenwood, C. R., Schulte, D., Arreaga-Mayer, C., & Terry, B. (1988). *Code for instructional structure and student academic response: Mainstream version* (MS-CISSAR). Kansas City: University of Kansas, Bureau of Child Research, Juniper Gardens Children's Project.

Castellan, N. J. (1979). The analysis of behavioral sequences. In R. B. Cairns (Ed.), *The analysis of social interactions: Methods, issues, and illustrations* (pp. 81–116). Hillsdale, NJ: Erlbaum.

Chard, D. J., & Kameenui, E. J. (2000). Struggling first-grade readers: The frequency and progress of their reading. *The Journal of Special Education, 34,* 28–38.

Deno, S. L. (1997). Whether thou goest . . . Perspectives on progress monitoring. In J. W. Lloyd, E. J. Kameenui, & D. Chard (Eds.), *Issues in educating students with disabilities* (pp. 77–99). Mahwah, NJ: Erlbaum.

Elbaum, B., Vaughn, S., Hughes, M. T., Moody, S. W., & Schumm, J. S. (2000). How reading outcomes for students with disabilities are related to instructional grouping formats: A meta-analytic review. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contempory special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 105–136). Mahwah, NJ: Erlbaum.

Fuchs, D., & Fuchs, L. S. (2001). One blueprint for bridging the gap: Project PROMISE: Practitioners and researchers orchestrating model interventions. *Teacher Education and Special Education, 24,* 304–314.

Fuchs, L. S., & Fuchs, D. (1986). Linking assessment to instructional intervention: An overview. *School Psychology Review, 15,* 318–323.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27–48.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children, 60,* 518–537.

Gersten, R., Morvant, M., & Brengelman, S. (1995). Close to the classroom is close to the bone: Coaching as a means to translate research into classroom practice. *Exceptional Children, 62,* 52–67.

Good, R. H., & Kaminski, R. (2000). *Dynamic indicators of basic early literacy skills* (DIBELS) [On-line]. Retrieved October 19, 2000, from http://dibels.uoregon.edu

Greenwood, C. R., & Abbott, M. (2001). The research to practice gap in special education. *Teacher Education and Special Education, 24,* 276–289.

Greenwood, C. R., Carta, J. J., Kamps, D., & Delquadri, J. (1993). *Ecobehavioral assessment systems software (EBASS): Practitioner's manual.* Kansas City: University of Kanas, Juniper Gardens Children's Project.

Greenwood, C. R., Delquadri, J., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology, 81,* 371–383.

Greenwood, C. R., Hart, B., Walker, D., & Risley, T. R. (1994). The opportunity to respond revisited: A behavioral theory of developmental retardation and its prevention. In R. Gardner, D. M. Sainato, J. O. Cooper, T. E. Heron, W. L. Heward, J. W. Eshleman, et al. (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 213–223). Pacific Grove, CA: Brooks/Cole.

Greenwood, C. R., & Maheady, L. (1997). Measurable change in student performance: Forgotten standard in teacher preparation? *Teacher Education and Special Education, 20,* 265–275.

Hollenbeck, A. R. (1978). Problems of reliability in observational research. In G. P. Sacket (Ed.), *Observing behavior: Data collection and analysis methods* (Vol. 2, pp. 46–62). Baltimore: University Park Press.

James, L., Abbott, M., & Greenwood, C. R. (2001). Writer's workshop, graphic organizers, and six-trait assessment: A winning writing strategy combo. *Teaching Exceptional Children, 33*(3), 30–37.

Juel, S., & Roper-Schneider, D. (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly, 20,* 143–152.

Kamps, D., Greenwood, C. R., & Leonard, B. (1991). Ecobehavioral assessment in classrooms serving children with autism and developmental disabilities. In R. J. Prinz (Ed.), *Advances in behavioral assessment of children and families* (pp. 203–237). New York: Jessica Kingsley.

Leibert, R. E. (1991). The Dolch list revisited: An analysis of pupil responses then and now. *Reading Horizons, 31,* 217–227.

Logan, K. R., Bakeman, R., & Keefe, E. B. (1997). Effects of instructional variables on engaged behavior of students with disabilities in general education classrooms. *Exceptional Children, 63,* 481–498.

Logan, K. R., & Stein, S. S. (2001). The research lead teacher model. *Teaching Exceptional Children, 33*(3), 10–15.

Lyon, R. G. (1999, October 26). *Education research: Is what we don't know hurting our children? Statement to the House Science Committee Subcommittee on Basic Research, U.S. House of Representatives.* Retrieved January 9, 2002, from http://www.nichd.nih.gov/crmc/cdb/r_house.htm

Mathes, P., Fuchs, D., Fuchs, L. S., Henley, A. M., & Sanders, A. (1994). Increasing strategic reading practice with Peabody classwide peer tutoring. *Learning Disabilities Research and Practice, 8,* 233–243.

O'Connor, R. E., Jenkins, J. R., & Slocum, T. A. (1995). Transfer among phonological tasks in kindergarten: Essential instructional content. *Journal of Educational Psychology, 87,* 202–217.

Pearson, P. T. (1987). *Silver Burdett Ginn: K–8 reading series.* Morristown, NJ: Author.

Perfetti, C. A. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33,* 283–319.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Raudenbush, S. W., & Chan, W. S. (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Consulting and Clinical Psychology, 61,* 941–951.

Rogosa, D. R., & Willett, B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrica, 50,* 203–228.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of research. *Review of Educational Research, 64,* 479–530.

Samuels, S. J. (1979). The method of repeated readings. *Reading Teacher, 32,* 4003–4408.

Scardamalia, M., & Breiter, C. (1986). Research on written comprehension. In M. C. Wittrock (Ed.), *Handbook of research in teaching* (Vol. 3, pp. 778–803). New York: Macmillan.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York: Guilford Press.

Simmons, D. C., Kuykendall, K., King, K., Cornachione, C., & Kameenui, E. J. (2000). Implementation of a schoolwide reading improvement model: "No one every told us it would be this hard!" *Learning Disabilities Research and Practice, 15,* 92–100.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Research Council, National Academy Press.

Taylor, B. M., Short, R. A., Frye, B., & Shearer, B. A. (1992). Classroom teachers prevent reading failure among low-achieving first-grade students. *The Reading Teacher, 45,* 592–598.

Teale, W. H., & Sulzby, E. (1987). Literacy acquisition in early childhood: The roles of access and mediation in storybook reading. In D. A. Wagner (Ed.), *The future of literacy in a changing world* (pp. 111–130). Elmsford, NY: Pergamon.

Vaughn, S., Hughes, M. T., Klingner, J., & Schumm, J. S. (1998). A collaborative effort to enhance reading and writing instruction in inclusion classrooms. *Learning Disability Quarterly, 21*(1), 57–74.

Walton, C. (1998). *The effects of professional development on change in instructional practice and student achievement: An experimental analysis.* Unpublished doctoral dissertation, University of Kansas, Lawrence.