Scaling Up an Early Reading Program: Relationships Among Teacher Support, Fidelity of Implementation, and Student Performance Across Different Sites and Years

Marc L. Stein
Vanderbilt University

Mark Berends
University of Notre Dame

Douglas Fuchs
Vanderbilt University

Kristen McMaster
University of Minnesota

Laura Sáenz
University of Texas Pan American

Loulee Yen, Lynn S. Fuchs, and Donald L. Compton
Vanderbilt University

Successful implementation of evidence-based educational practices at scale is of great importance but has presented significant challenges. In this article, the authors address the following questions: How does the level of on-site technical assistance affect student outcomes? Do teachers' fidelity of treatment implementation and their perceptions of school climate mediate effects on student performance? Using a randomized control trial at scale, the authors examine Kindergarten Peer Assisted Learning Strategies, which previously has been shown to be effective in increasing student reading achievement. Analyzing data from 2 years and three sites, the analyses show that the level of on-site technical support has significant effects on reading achievement gains, are robust across multiple sites, and are mediated by fidelity of implementation within teachers' classrooms.

Keywords: *early reading achievement, implementation fidelity, randomized control trial, peer-assisted learning strategies*

Bridging research and practice is a persistent problem in education and a perennial impediment to the implementation of educational reforms. It is widely acknowledged by researchers and policy makers that many teachers are slow to adopt instructional innovations meant to modify core teaching practices (Berends, 2004; Elmore, 1996, 2004; Glennan, Bodilly, Galegher, & Kerr, 2004; Schneider & McDonald, 2007). For example, teachers may implement instructional reforms during a study and then, slowly or immediately, partially or completely, revert back to

conventional modes of teaching. Such poorly sustained instructional implementation may be observed even when the instructional reform is evidence based and can lay claim to improving student outcomes.

In this era of No Child Left Behind, the federal government has placed strong emphasis on the use of evidence-based practices, on increasing academic achievement across traditionally enfranchised and disenfranchised student groups, and on holding schools and teachers accountable for their students' achievement. As accountability increases, the demand for evidence-based interventions and reforms is also likely to increase. In this policy environment, it becomes especially important that we understand factors mediating teachers' implementation of reforms so that facilitating factors may become part of the educational interventions and implementation process. Because increasing student achievement is the ultimate goal, it is also necessary to understand how these mediating factors interact with and affect student achievement. Considerable resources are spent every year at all levels in the education system to improve student outcomes. If teachers cannot access and engage in instructional innovations—if they continue to rely on outdated modes of instruction—then these expenditures and efforts have been wasted. As Berman and McLaughlin (1976) stated, "The bridge between a promising idea and its impact on students is implementation . . . [however] innovations are seldom implemented as planned" (p. 349).

Not only is there demand for research-based practices but a demand, too, that these practices are brought to scale to affect large numbers of schools, teachers, and students. By better understanding the factors that promote teacher implementation of interventions at scale, program developers and researchers can create innovations that will more likely affect lasting change and widespread student achievement. Schneider and McDonald (2007) write,

> Scale-up research is translational research. It is conducted with the explicit objective of informing practice—which means not only documenting the importance of implementing interventions with integrity, but documenting the benefits of balancing fidelity of implementation with adaptation to dynamic local contexts. (p. 11)

It is important to understand the factors related to variation among teachers, schools, and fidelity of implementation and performance and also how these factors play out when an intervention is brought to scale (Dusenbury, Brannigan, Falco, & Hansen, 2003).

The purpose of this study was to look at these issues by examining a specific evidence-based intervention, Kindergarten Peer Assisted Learning Strategies (K-PALS), which is currently being studied at scale. Specifically, we address the following research questions:

- What levels of technical assistance, or support, are necessary to ensure strong teacher implementation of K-PALS?
- Do the levels of teacher support differentially affect K-PALS implementation?
- Do teacher characteristics and other factors mediate the implementation of this intervention?
- Do teacher perceptions of their school context also play a mediating role?
- How does teacher implementation of K-PALS affect student achievement?

Within the context of scaling-up a research-based instructional intervention, we hope our analyses shed light on the effects of implementation on student achievement and the possible importance of other factors, providing lessons helpful to policy and practice in the area of early reading achievement.

We begin with a brief description of the K-PALS program and our current study of it at scale. Next, we describe the theoretical framework that provides a rationale for our study. Then, we discuss the nature of our data and our

analyses. Finally, we present our results and discuss their implications.

Overview of the K-PALS Program

The original PALS program was designed to help teachers in Grades 2 through 6 differentiate instruction for students across a wide range of achievement levels as well as to modify instruction for students with disabilities (Fuchs & Fuchs, 1998; Fuchs, Fuchs, Al Otaiba, et al., 2001; Fuchs, Fuchs, Mathes, & Simmons, 1997; Fuchs, Fuchs, Thompson, Yen, et al., 2001). PALS was created to supplement rather than supplant reading programs already implemented in classrooms to better align with the existing programs as teachers attempted to meet accountability demands. K-PALS is a downward extension of the original Grades 2 through 6 PALS program, with appropriate modifications of program features to make it more age appropriate.

PALS represents a modest imposition of time on the classroom teacher, as it was designed to be implemented three times a week for approximately 35 min a session. It calls for the pairing of all students in a classroom, whereby one member is a stronger reader and the second is a weaker reader. Pairs work through structured activities while the classroom teacher monitors and provides feedback. The goal of the program is that stronger students will help struggling students to become better readers through direct interaction and tutoring. The process has also been shown to increase the skills of the stronger students. PALS also includes a motivational component, as the student pairs are rewarded with points based on their progress as a team through the activities (Fuchs, Fuchs, Al Otaiba, et al., 2001).

In several randomized field trials across different grade levels in numerous schools in Nashville, Tennessee, PALS has been empirically shown to increase student reading performance. For example, in one school in which teachers were implementing PALS in first and second grades, students substantially improved their scores on the statewide reading test, with the typical score increasing from the 28th to the 52nd percentile (Fuchs, Fuchs, Al Otaiba, et al., 2001). Other studies have documented that the program promotes reading gains for low and average students as well as for students with learning disabilities and behavior problems (Barton-Arwood, Falk, & Wehby, 2005; Fuchs et al., 1997; Fuchs, Fuchs, Thompson, Al Otaiba, et al., 2001; Fuchs, Fuchs, Thompson, Yen, et al., 2001). In the 1990s, PALS-Reading (and PALS-Math) were approved by the U.S. Department of Education's Program Effectiveness Panel for inclusion in the National Diffusion Network on effective educational practices. Recently, the U.S. Department of Education's What Works Clearinghouse found PALS "to have potentially positive effects on reading achievement" for English language learners and non–English language learners.

We and our colleagues are currently funded by a 5-year grant from the Institute of Education Sciences of the U.S. Department of Education to study the scale-up of PAL using a multisite, randomized control trial. In addition to conducting this research in the Metropolitan Nashville Public Schools, we are working with teachers in several urban districts in Minnesota (Minneapolis, St. Paul, and Bloomington) and in schools in six contiguous districts in South Texas. In the first 2 years of this research project, we recruited Title I and non–Title I schools at each of the three sites, and within schools, kindergarten teachers were randomly assigned to one of four study conditions: (a) controls who by definition did not participate in K-PALS, (b) 1-day workshop before K-PALS implementation, (c) workshop plus a "booster" training workshop, or (d) workshop, booster, and a classroom helper (a research assistant knowledgeable in K-PALS).

Fidelity of Implementation: Program Characteristics and Setting Context

Prior research on the implementation of educational interventions provides a useful framework with which to explore mediating factors of teachers' fidelity of implementation. It is posited that there are two main dimensions across which fidelity of implementation can vary: (a) program characteristics and (b) features of settings into which the programs are placed (see Glennan, Bodilly, Galegher, & Kerr, 2004; Ruiz-Primo, 2005). Figure 1 portrays these two dimensions with the characteristics of the program intervention influencing the fidelity of implementation. Educational interventions
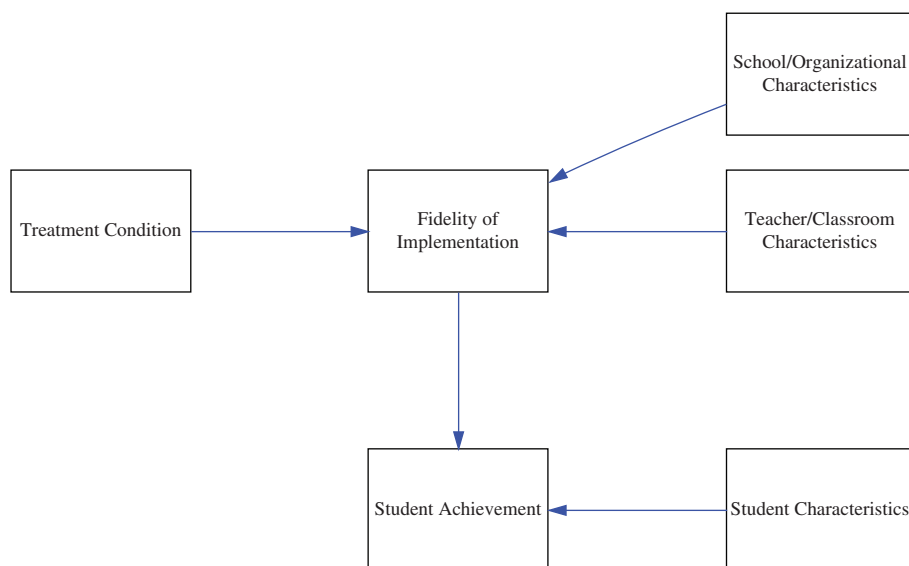
FIGURE 1. *Importance of program characteristics and setting to fidelity of implementation.*

are implemented within a social context that includes, among other things, the varying conditions of school organizations, teachers, and classrooms. These conditions also influence the degree of implementation fidelity, which in turn affects student achievement outcomes. The figure also portrays that student characteristics (e.g., socioeconomic status, race–ethnicity, gender, special education designation, and English language learner status) may also be related to student achievement, even within a randomized control trial where the intervention is randomly assigned to teachers. We discuss program characteristics and setting context further in the sections that follow.

Program characteristics

Research has shown that several factors related to the program itself are associated with higher levels of fidelity of implementation, such as specific materials to support implementation, a targeted focus of the intervention, and training and supportive professional development of teachers (Glennan et al., 2004; Moncher & Prinz, 1991). These factors are very much part of the K-PALS program and are present in our three treatment conditions. In K-PALS, same-age children

work in dyads through 72 structured lessons. Teachers follow their judgment in creating these dyads: They pair their strongest learners with their weakest learners, the second-strongest child with the next-to-weakest child, and so on. If a teacher believes a resulting pair is socially incompatible, the pair is reassigned. Each student in a pair takes a turn as a reader (tutee) and coach (tutor); that is, roles are reciprocal. Pairs remain together for 4 to 6 weeks, at which point the teacher forms new groups. Teachers encourage their students to work productively and cooperatively. Following 3 weeks of training, K-PALS was conducted by study teachers four times per week in 35-min sessions for 18 weeks.

For each of the 72 peer-mediated lessons, children engage in the same four activities: What Sound (learning letter–sound correspondence), Sound Boxes (learning decodable words), Sight Words (learning sight words), and Reading Sentences.

What Sound consists of a series of letters displayed in a left-to-right, line-by-line format. In Lessons 1 through 9, there is an average of 15 letters; in Lessons 10 through 21, 18 letters; and in Lessons 22 through 72, 24 letters. All letter sounds are introduced by Lesson 50. Digraphs (*sh*, *ch*, *th*, *ck*) are first presented in Lesson 55.

New sounds are introduced at the beginning of the lesson, enclosed by a box and next to a picture illustrating that sound (e.g., picture of an apple to the right of the letter *a*).

The coach points to a letter and prompts the reader to say its sound by saying, "What sound?" When coaches come upon the shape of a star interspersed among the letters, they praise their reader (e.g., "Great job!"). If the reader makes a mistake or does not know the sound of a letter, the coach stops the reader, tells the correct sound, asks the reader to repeat the correct sound, then directs him or her to read the line again (i.e., "Stop. That sound is __. What sound? Good. Go back and read that line again"). When the reader has said all of the sounds, the coach draws a line through a happy face on the bottom of the lesson sheet. The partners switch roles and repeat the activity. What Word, Sound Boxes, and Read the Sentence activities follow a very similar procedure.

K-PALS provides a manual to all teachers that explicitly and thoroughly explains the program and provides guidelines for its implementation. K-PALS teachers are also given a variety of additional materials (e.g., all required worksheets, student rewards, folders, and score sheets) to help them with program implementation. Because of this, they do not have to create their own materials, and hence, we would expect higher levels of fidelity of implementation. Thus, these K-PALS features, too, would seem to strengthen the probability of higher fidelity of implementation.

Given the expected high fidelity of K-PALS implementation across our study teachers, we were interested in whether different levels of technical assistance, or teacher support, would engender greater levels of fidelity of implementation. We conceptualized teacher support in terms of a combination of two factors identified in the literature as related to higher fidelity: initial training and supervision (Moncher & Prinz, 1991; Ruiz-Primo, 2005). In their review of the experiences of several educational model developers, Glennan et al. (2004) stated that the intervention developers

> often pointed to the difficulties of changing practice and reported adding new forms of support and refining those already in place to ensure that school staff understood the reasons for change, understood why particular practices were being emphasized, understood and became adept at new practices, and felt supported in the long term. (p. 655)

Following Glennan et al.'s review, we aimed to understand whether different levels of teacher support directly influence the K-PALS teachers' implementation and indirectly affect the reading performance of children in schools in Nashville, Minnesota, and South Texas.

As indicated, we randomly assigned teachers to one of four study groups, with each successive condition representing an increase in the level of support teachers received.

1. Control group. Teachers received no training and did not implement the K-PALS program.
2. Workshop group. Teachers attended a pre-implementation day-long training workshop during which we explained previous research indicating the positive effects of K-PALS on student reading achievement. Teachers viewed videos of K-PALS lessons implemented in urban and suburban classrooms. K-PALS staff also taught teachers key program components and required them to implement specific lessons through role-play activities.
3. Booster group. Teachers received the initial workshop as well as participated in two follow-up booster sessions. The purpose of the booster sessions was to provide teachers with opportunities to review K-PALS procedures, to identify implementation issues, and to problem solve with other teachers and the researchers.
4. Helper group. Teachers in the helper group participated in the initial training workshop and follow-up booster sessions. In addition, a trained graduate assistant (i.e., helper) provided weekly technical assistance as they implemented K-PALS in their classes.

Setting context

In addition to the importance of program characteristics, theory and research on implementation of prevention programs in education settings and teacher change in response to professional development indicate the importance of the social context in which educational interventions are implemented (Dusenbury et al., 2003; Ruiz-Primo, 2005; Smylie, 1988). Additionally, research suggests that teacher

and school characteristics have mediating effects on quality of program implementation and educational outcomes.

Prior research on fidelity of program implementation leads us to expect that fidelity will vary with certain teacher characteristics. Teachers are the "street level bureaucrats" at the core of educational change (Weatherly & Lipsky, 1977). Without the dedication of teachers who embrace the educational intervention, no reform will be enacted, no matter how effective it may be (Berends, Bodilly, & Kirby, 2002; Fullan, 2001). Teachers carry with them a great deal of knowledge based on their educational attainment, teaching experience, and other personal characteristics that together are likely to be related to their engagement in whole-school restructuring activities (Louis & Marks, 1998). Thus, it is important to examine the relationships among various teacher background characteristics, implementation of educational interventions, and effects of the program on student outcomes.

For example, consider the following teacher characteristics that have been shown or hypothesized in the literature to affect fidelity of implementation, such as teacher experience and self-perceived efficacy. Rohrbach, Graham, and Hansen (1993) state that new and inexperienced teachers are more likely to continue use of a program than experienced teachers. Gersten, Chard, and Baker (2000), citing Huberman (1995), note that teachers in the early stages of their careers are characterized by a "survival" and "exploration" orientation as they search for instructional methods and materials to guide their instruction. "Sometimes this survival mode results in greater openness to trying new approaches in the hope that something will help stabilize their instruction" (Gersten et al., p. 452). By contrast, it is likely that more experienced teachers have discovered instructional strategies that work for them, have built up a repertoire of strategies that they use regularly, and consequently, are more resistant to change (see Coburn, 2004). Thus, we might expect that newer teachers will be more likely to have higher fidelity of implementation than their more experienced peers.

In addition, it is important to examine teacher efficacy as an important social-context measure, because previous research has also shown that teachers' self-perceived efficacy is related to their implementation of instructional innovations (Dusenbury et al., 2003; Moncher & Prinz, 1991; Ruiz-Primo, 2005; Smylie, 1988; Vaughn, Klingner, & Hughes, 2000). Teachers who view themselves as more efficacious with students may also be likely to implement scientifically based educational interventions.

The social context of schools is also likely to influence the fidelity of implementation and program effects. Research also indicates that school-level characteristics are related to teacher fidelity of implementation of educational interventions and student achievement. These include (a) the instructional leadership of the principal and (b) school climate, staff morale, and communication within the school community (Dusenbury et al., 2003; Fullan & Pomfret, 1977; Gottfredson, 1984). Both of these characteristics—leadership and community—are important influences on the willingness and ability of teachers to implement new educational programs.

Research has consistently shown that the principal strongly influences the likelihood of program implementation and change (Berends et al., 2002; Fullan, 2001; Berman & McLaughlin, 1978). Leadership of the principal may translate into the ability to encourage teachers to implement programs and obtain sufficient resources for teachers in their efforts to implement change.

Implementing educational interventions may also result in greater fidelity of implementation if the program is aligned with other school curricula and instructional programs (Newmann, Smith, Allensworth, & Bryk, 2001) implemented by a community of teachers who work together frequently outside of their classrooms to focus on the academic needs of students (Louis & Marks, 1998). Thus, we examine in our analyses how implementation fidelity of K-PALS is related to these school context features.

It is important to note that with randomization in this K-PALS study at the level of the teacher (i.e., teachers randomly assigned to conditions), our a priori expectation is that effects of teacher-level characteristics will be controlled through randomized assignment. Students, however, may not have been randomly assigned to teachers' classrooms. Thus, we may expect students to influence teachers' perceptions, especially with regard to teachers' perception of their efficaciousness, or ability to teach their students.

Method

Sites, Schools, Teachers, and Students

Sites

Three sites participated in this study: Nashville, Minnesota, and South Texas. In the Metropolitan Nashville Public Schools, where we have conducted much of our previous PALS research, the 66 elementary schools are divided evenly between Title I and non–Title I. Non-Hispanic Whites, African Americans, Hispanics, and "Others" constitute 37%, 42%, 18%, and 3%, respectively, of the student community. Sixteen percent of students receive special education services. As described, the Minnesota site includes the school districts of Minneapolis, St. Paul, and Bloomington. The Minneapolis Public Schools, from which most of the Minnesota sample came, serve an elementary school population that is 45% African American, 14% Asian, 11% Hispanic, 4% Native American, and 26% non-Hispanic White. More than half of the students (57%) in Minneapolis receive subsidized lunch, 14% get special education services, and 24% are English language learners, a group representing 80 home languages. The South Texas site involves six school districts in Hidalgo County, just north of the border with Mexico, which ranks among the poorest counties in the nation, with 36% of its population living in poverty. The six school districts collectively have 45 elementary schools serving a population that is 80% Hispanic, of which 99% are Spanish speaking, with many considered limited English proficient.

School selection

At each of these three sites, we recruited schools in Years 1 and 2 by blocking on whether they were Title I or non–Title I (Nashville), whether they were Title I or non–Title I and offered full-day or half-day kindergarten (Minnesota), and whether they had high or low proportions of limited-English-proficiency children (South Texas). Project staff at each site discussed the study with building principals and teachers. In Year 1, 48 schools agreed to participate in the study (10 in Nashville, 19 in Minnesota, and 17 in South Texas), and in Year 2, 49 schools participated (14 in Nashville, 21 in Minnesota, and 14 in South Texas). Across Years 1 and 2, there were 71 total schools that participated in the three sites (14 in Nashville, 36 in Minnesota, and 21 in South Texas).

Teacher selection

Once schools agreed to participate, project staff met with teachers to explain study participation, including the necessity for random assignment to treatment and control conditions. We recruited 145 teachers in Year 1 (52 in Nashville, 42 in Minnesota, and 51 in South Texas) and 134 teachers in Year 2 (54 in Nashville, 40 in Minnesota, and 40 in South Texas), who were assigned randomly within schools to either the control group or to one of the three treatment conditions (workshop, workshop and booster, or workshop and booster and helper). Teachers were given a modest cash stipend in return for their study participation and all that it entailed, including their willingness to complete surveys and participate in structured interviews. Of the 134 teachers who participated in Year 2 of the study, 55 were also a part of the 1st year of the study. Therefore, the total number of unique teacher participants across the 2 years of the study was 224, with an additional 55 teachers receiving a "double dose" of K-PALS because of their participation across the 2 years.

Student selection

In each study classroom, we distributed parent consent forms to all students, printed in English (and Spanish, Somali, and Hmong, where appropriate). We administered reading tests to all kindergartners for whom we obtained parental consent (more than 90% across sites). These tests were Rapid Letter Naming and Rapid Letter Sounds (RLS). On the basis of students' performance on these measures, we identified 4 children with the lowest reading scores, 4 students with scores in the middle of the distribution, and 4 students with the highest scores within each class. We checked the validity of these grouping with classroom teachers. If there was disagreement between the students' scores and teacher judgments, the scores were given greater weight in determining the students' final status. We also included all children with identified disabilities. Pretreatment and posttreatment tests were individually administered to 1,674 kindergarten

students in Year 1 and 1,555 in Year 2, for a total of 3,229 students in the study across the 2 years.

Staff and classroom-based assistance

Each site had its own staff responsible for all project-related activities in Years 1 and 2. In Nashville, there was a principal investigator (and overall project director), two full-time project coordinators, and seven research assistants. One project coordinator was responsible for day-to-day management of the study in Nashville and for coordinating study activity across the three sites. In Minnesota, there was a principal investigator and eight research assistants. In South Texas, there was a principal investigator and seven research assistants. Research assistants at each of the sites were graduate students who typically worked 10 to 20 hr per week. Their responsibilities included administering pre- and posttreatment measures, gathering student and teacher demographic data, and assisting at least 1 to 4 K-PALS teachers (in the *helper* condition). In Nashville only, staff were responsible for double-scoring tests, entering data from all three sites into an electronic database. All data were double-scored in Nashville. All double-scored data were then entered twice into separate databases and went through a series of electronic and hand checks by a data manager.

Measures

Student reading achievement

Pretreatment measures were administered to all students in the sample approximately 3 weeks prior to the first K-PALS workshop. Posttreatment measures were administered approximately 20 weeks after the administration of pretests. Student reading achievement gains were calculated as the pre- to posttreatment gain on the RLS test. The same measures administered in Year 1 were again administered in Year 2.

The RLS is based on a measure developed by Levy and Lysunchuk (1997) to assess the number of letter sounds a student can identify in 1 min. The student is presented with a sheet containing four practice letters (*b*, *c*, *h*, *a*—which also appeared in the test battery) and a test battery of all 26 lowercase letters in a random order. First, the examiner and child engage in several practice

items to ensure that the child understands to provide letter sounds, not letter names. The examiner says, "I'm going to show you some letters. You tell me what sound the letters make. If you don't know the sound a letter makes, don't worry. What's important is that you try your best." The examiner proceeds by showing the student the practice letters, saying, "This letter says /b/. Your turn. What sound does it say?"

Next, the examiner uncovers the first line of letters of the test and says,

> You're doing a great job. Now it's just going to be your turn. Go as quickly and carefully as you can. Remember to tell me the sounds the letters make. Try your best. If you don't know a letter sound, it's okay.

If the student does not provide a response within 3 seconds, the examiner prompts the student to go on to the next sound. The test is stopped at 1 minute. Only one sound is considered as the correct response for each letter. For example, only short vowel sounds are accepted as correct. For *g*, only the hard sound heard in *gate* is correct; the soft sound as heard in *cage* is incorrect. The Spanish sound for *j* (sounds like /h/) is also not accepted. Sounds are accepted as correct if they are followed with the "uh" sound (i.e., *buh* for /b/, *cuh* for /c/, etc.).

The number of sounds named correctly is the RLS score. If the child completes the 26-sound battery prior to 1 minute, the amount of time is noted and the test is stopped. To prorate such scores, the number of sounds named correctly is divided by the amount of time (i.e., in seconds) taken to complete the test. This number is then multiplied by 60 to obtain the adjusted RLS score. A prorated score was calculated for a very few children. The RLS test was administered at pre- and posttreatment (see descriptive statistics of this measure in Table 1).

We consider the RLS as an appropriate indirect measure of K-PALS implementation because it measures knowledge (letter–sound correspondence) that is an important part of the K-PALS curriculum. It is important in two ways. First, RLS is an ever-present component in each of the 72 K-PALS lessons; that is, young children are expected to master a set of increasingly sophisticated letter sounds across the lessons. Second, the letter sounds represent a pivotal building block for

TABLE 1
Rapid Letter Sounds Pre- and Posttest Means, Standard Deviations by Treatment Condition and Site

Site	Control		Workshop		Booster		Helper		Site total	
	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest	Pretest	Posttest
Nashville										
M	12.80	29.40	11.30	39.70	11.40	42.70	11.90	42.80	11.80	39.30
SD	(11.40)	(15.30)	(11.70)	(18.80)	(11.40)	(16.90)	(12.60)	(20.30)	(11.80)	(18.80)
n	249	249	323	323	310	310	350	350	1,232	1,232
Minnesota										
M	11.50	35.00	11.90	41.30	13.90	43.10	13.60	38.10	12.70	40.20
SD	(11.50)	(17.70)	(12.70)	(20.60)	(13.10)	(18.90)	(12.10)	(17.20)	(12.60)	(19.30)
n	179	179	312	312	294	294	114	114	899	899
South Texas										
M	17.80	35.00	16.50	37.50	15.00	39.60	18.70	37.80	16.60	37.60
SD	(15.10)	(17.60)	(15.60)	(17.00)	(14.10)	(17.50)	(15.20)	(17.00)	(15.00)	(17.40)
n	240	240	333	333	327	327	140	140	1,040	1,040
All sites										
M	14.20	32.90	13.30	39.50	13.50	41.70	13.80	40.80	13.60	39.00
SD	(13.10)	(17.00)	(13.70)	(18.90)	(13.00)	(17.80)	(13.40)	(19.10)	(13.30)	(18.50)
n	668	668	968	968	931	931	604	604	3,171	3,171

+	−	NA		Teacher-Directed What Sound Checklist
				Time Started:
			1	Teacher introduces the *new* sound.
			2	Students repeat the new sound (check "NA" if there is not a new sound today).
			3	Completes entire What Sound? activity with class at least one time.
			4	Uses appropriate prompt, "What Sound?"
			5	Says something nice at each star.
			6	If demonstrates correction procedure, uses exact/consistent wording on bolded portions and incorporates all 5 steps of the correction procedure.
			a	*The Coach tells the sound (e.g., "Stop.* **That sound is ____. ")**
			b	*The Coach tells the Reader to repeat the sound (e.g., "* **What sound? ")**
			c	*The Reader* **repeats the sound**.
			d	*The Coach* **asks** *the Reader* **to repeat the line** *(e.g., "Good. Go back and read that line again.")*
			e	*The Reader* **reads the line again**.
			7	Teacher calls on students to respond (i.e., individual or whole class responses are both acceptable).
			a	Calls on whole class to respond for some items.
			b	Calls on groups of students to respond for some items (i.e., all girls, all boys).
			c	Calls on individual students to respond for some items.
			8	Students respond with appropriate response.
			9	Most students are responding during choral responses.
			10	Most students are paying attention while others make individual responses. Includes when an individual student is the Coach.
			11	Most students are on-task during the What Sound activity.
			12	Teacher provides students with help when needed. Includes assistance with prompts, procedures, etc., as well as helping students with the correction procedure.
			13	Students listen to correction and repeat correct response. Does not include accepting help on prompts and procedures. This item only applies to the correction procedure.
				How many times did the teacher go through the What Sound activity?
				Time Ended:
A = Total (+)	B= Total (-)	C = Total (+) and (-)	A/C * 100 = % accuracy	Comments:
Teacher-Directed What Sound Fidelity				

FIGURE 2. *Teacher-Directed What Sound Checklist.*

the teaching of decodable words—another important focus of the K-PALS program. Thus, strong gains from pretreatment to posttreatment on the RLS measure is consonant with a view that K-PALS has been properly implemented.

Prior to pretreatment testing, the test trainers (i.e., the project coordinators in Nashville and coprincipal investigators in Minnesota and South Texas) introduced testing procedures and administration instructions. At this meeting, the project coordinator sat in via teleconference to ensure similarity in the training and that everyone was on the same page regarding the measures and student handling. Research assistants practiced administering the reading tests with each other. At each site, after research assistants completed substantial practice sessions with a preselected partner (usually experienced tester with new tester), the test trainer and research assistants established interrater agreement on administration and scoring procedures for all

measures. These training activities occurred both at pre- and posttreatment testing. Interrater agreement was established each year at pre- and posttreatment. In Nashville and Minnesota, interrater agreement was 96%. In South Texas, pre- and posttreatment reliability was 99% across Years 1 and 2.

Fidelity of implementation. As part of this investigation, the study team designed rubrics to assess teachers' fidelity of K-PALS implementation for each of the major components (see Figure 2 for an example of the "Teacher-Directed What Sound" fidelity rubric). The fidelity checklist had 68 teacher items and 122 student items. The majority of the student items focused on their ability to follow the implementation procedures for the different activities. There were also items addressing organization (i.e., getting their folders and materials ready quickly) and compliance with K-PALS rules (i.e., cooperating, being on task). During fidelity, the teacher and the student pairs were observed. One student pair was observed for each peer-mediated activity. Each item on the checklist was marked as observed (+), not observed (−), or not applicable (NA). Whereas in the *helper* group condition, graduate students assisted the teacher in the training of students, the teacher's implementation of K-PALS (never the graduate student's implementation) was observed to determine fidelity of treatment implementation.

There were two rounds of fidelity checks in each study classroom during the 20-week implementation. The two project coordinators from Nashville conducted these fidelity observations across all three sites. For the first fidelity check, the project coordinators conducted six observations together to establish interrater agreement, which was well above 90%. For the second fidelity check, the same two project coordinators conducted five observations together to establish interrater agreement, which was also greater than 90%.

A teacher's total fidelity of implementation was the percentage of items on the scoring rubrics across all K-PALS components rated as observed by the project coordinator. For purposes of this study, a measure of average implementation fidelity was created by taking the average of the two individual fidelity scores.

Measures of teachers' perception of school context

From the pretreatment teacher surveys, we constructed scales that measure teachers' perceptions of the school contexts in which they worked. For all items in these scales, discussed below, teachers were asked to respond *strongly disagree* (1) to *strongly agree* (4). Table 2 provides descriptive statistics for these scales.

Instructional Coherence is a scale composed of eight teacher survey items ($\alpha = .80$) that measures the extent to which a teacher feels that the curriculum, instruction, and learning materials are consistent both across and within grade levels; the extent to which the teacher believes that there is continuity in programs implemented in the school; and the extent to which changes introduced at the school help to promote the school's learning goals. For the entire sample, the mean instructional coherence was 2.71, with a standard deviation of 0.37.

Teacher Community is a scale of five items ($\alpha = .85$) that measures the extent to which a teacher feels that his or her colleagues share a focus on student learning and beliefs and values in what the central mission of the school should be, the extent to which the school has clear strategies for instruction, and the extent that teacher morale is high. The mean for teacher community was 3.15 (*SD* of 0.52).

Principal Leadership is a scale composed of 15 items ($\alpha = .93$) that measures the extent to which the teacher perceives that the principal of the school communicates clear expectations to, is supportive of, and is respecting of teachers as educators as well as items that gauge the extent that the principal is an instructional leader in the school. The average in the sample for this measure was 2.97, with an *SD* of 0.53.

Teacher Efficacy is measured by three items, which did not exhibit an acceptable alpha reliability ($\alpha = .30$) to warrant their combination into one scale measuring teacher efficacy. Therefore, they were treated as separate items. The first, Efforts and Ability, asks teachers the extent to which their success or failure is due primarily to factors beyond their control rather than their efforts and ability (mean of 2.63 and *SD* of 0.73). The second,

TABLE 2

Student and Teacher Level Descriptive Statistics

Variable	*M*	*SD*
Student-level variables		
Pre- to Posttest Gain		
Rapid Letter Sound (RLS)	27.16	15.87
English language learner	0.24	0.43
Free or reduced lunch eligible	0.62	0.49
African American	0.25	0.43
Non-Hispanic White	0.26	0.44
Hispanic	0.40	0.49
Asian	0.05	0.22
Other	0.03	0.16
Female	0.50	0.50
Individualized Education Plan	0.05	0.22
Weeks between testing times	25.78	2.93
Teacher-level variables		
Average fidelity (whole sample)	85.71	11.54
Workshop	80.72	12.21
Booster	87.38	10.34
Helper	91.37	8.64
Efforts and ability	2.63	0.73
Attitudes and habits	2.52	0.66
School resources	3.01	0.56
Teacher community	3.15	0.52
Instructional coherence	2.71	0.37
Principal leadership	2.97	0.53
K-PALS instructional coherence	2.85	0.36
K-PALS principal leadership	2.22	0.76
Second year in study	0.13	0.34

Note. K-PALS = Kindergarten Peer Assisted Learning Strategies.

Attitudes and Habits, explores the extent to which teachers see their students' attitudes and habits as reducing their chances of successfully teaching them (mean of 2.52 and *SD* of 0.66). Finally, School Resources asks the extent to which teachers believe students can learn with the school resources available to them. This item was reverse coded to maintain the same directionality in meaning of responses (mean of 3.01 and *SD* of 0.56).

K-PALS Instructional Coherence is a scale composed of three items ($\alpha = .63$) gauging the degree to which teachers believe that K-PALS fits with the current instructional program and goals of the school. The mean for this measure was 2.85, with an *SD* of 0.36.

K-PALS Principal Leadership is a scale consisting of two items ($\alpha = .84$) that measures the extent to which the principal talks frequently with the teacher about K-PALS and takes a personal interest in ongoing K-PALS assistance for teachers (mean of 2.22 and *SD* of 0.76).

Student-level measures

Our analyses also include several student-level measures. We included dummy variables for whether the student was an English language learner, received free or reduced-price lunch, or had an individualized education plan. We also included dummies for race–ethnicity (White as reference group) and gender (female = 1). The descriptive statistics for these measures appear in Table 2.

Because we wanted complete information for all the measures listed above for students, teachers, and schools, the final sample for analyses

consisted of 2,959 students and their 259 teachers, nested within 67 schools.

Analytic Approach

Effects of level of teacher support and fidelity on student achievement

Noting the hierarchical nature of theories of teacher change and implementation of research-based practices as well as the inherent hierarchical structure of educational settings with students nested in teachers' classrooms nested in schools, we employed a hierarchical linear modeling strategy to investigate the effects of the level of support provided to teachers and the effect of the level of teacher fidelity of implementation of the K-PALS reading program on student reading achievement gains.

Consider the three-level hierarchical model

$$\Delta Y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{Minority})_{ijk} + \pi_{2jk}(\text{ELL})_{ijk} + \pi_{4jk}(\text{FRL})_{ijk} + \pi_{5jk}(\text{IEP})_{ijk} + \pi_{6jk}(\text{Weeks})_{ijk} + e_{ijk}$$

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}\text{Workshop}_{jk} + \beta_{02k}\text{Booster}_{jk} + \beta_{03k}\text{Helper}_{jk} + r_{0jk}$$

$$\beta_{00k} = \gamma_{000} + \gamma_{001}\text{Minn} + \gamma_{002}\text{SouthTexas} + \gamma_{003}\text{Year} + u_{00k}, \qquad (1)$$

where

- ΔY_{ijk} is the achievement gain, pre- to posttest, of student i in classroom j and school k;
- π_{0jk} is the mean achievement of classroom/teacher j in school k adjusted for student demographics;
- e_{ijk} is a random student effect—the deviation of student ijk's score from the classroom mean, assumed to be normally distributed with mean of 0 and variance of σ^2;
- Minority is a set of dummy variables for students' race–ethnicity (Black, Hispanic, Asian, and Other), with White students as the reference group;
- ELL is a dummy variable coded 1 if the student is categorized as an English language learner;
- FRL is a dummy variable coded as 1 if the student is eligible for free and reduced-price school lunches;
- special education status of individual students is represented by the dummy variable IEP, which takes the value of 1 in the case of

a student with an individualized education plan; and finally,
- Weeks represents the amount of time, in weeks, that elapsed between the pre- and posttests used to compute the individual reading test gain scores.

All of the student demographic dummy variables were entered into the model uncentered, and the variable weeks was entered grand mean centered to adjust for the average time elapsed between pre- and posttreatment testing of the sample.

At Level 2, the teacher level, the adjusted student-level intercept, π_{0jk}, was modeled with the inclusion of dummy variables indicating the level of support received by the individual teachers—Workshop, Booster, and Helper—with the control teachers serving as the reference group. The intercept was allowed to vary randomly across level two where r_{0jk} was a random classroom or teacher effect—the deviation of classroom or teacher jk's score from the school mean, assumed to be normally distributed with mean of 0 and variance of τ_π. All other Level 1 variables were fixed and not permitted to vary randomly across classrooms, as the effect of student demographics was modeled as being constant across classrooms.

At Level 3, we entered dummy variables for the two sites other than Nashville that are a part of the study, Minnesota and South Texas. We also included a dummy variable for the year of the study (Year 1 and Year 2) that took on the value of 1 if teachers were in the 2nd year. A random school effect (u_{00k}) was the deviation of school k's score from the grand mean. This effect was assumed to be normally distributed with mean of 0 and variance of τ_β.

To capture the possible mediating effect of average fidelity of implementation on the effect of level of teacher support on reading achievement, a measure of the individual teacher's average fidelity of implementation was added to Equation (1) such that Level 2 became

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}\text{Workshop}_{jk} + \beta_{02k}\text{Booster}_{jk} + \beta_{03k}\text{Helper}_{jk} + \beta_{04k}\text{avgfid}_{jk} + r_{0jk}. \qquad (2)$$

If the effect of level of teacher support on student achievement is mediated by the average fidelity with which the teacher implements the K-PALS

program, we would expect that the estimated coefficients on the teacher support–level dummies would be adjusted toward zero and that they would become statistically insignificant by the inclusion of the average fidelity measure, which should show a statistical significant effect in Equation (2).

Mediating effects on average fidelity of implementation

To investigate the effects of teachers' perceptions of their school contexts on average fidelity of implementation, we estimated the following two-level hierarchical models where the sample has been restricted to only those teachers who were implementing the K-PALS program (i.e., control teachers were excluded from the analysis):

$$\text{avgfid}_{ij} = \beta_{0j} + G_{ij}\beta_{1j} + r_{ij0}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}. \qquad (3)$$

avgfid represents the average fidelity of implementation of teacher i in school j; β_{0j} is the mean average fidelity in school j adjusted for the treatment group, G_{ij} a vector of two dummy variables for the booster and helper groups with the workshop only group as the comparison, and r_{ij} is a random teacher effect—the deviation of teacher ij's score from the school mean, assumed to be normally distributed with mean of 0 and variance of σ^2. The intercept β_{0j} was allowed to vary randomly at Level 2. This model provided us an estimate of the effects of levels of teacher support on teacher fidelity of implementation scores unadjusted for site and teacher perceptions of school context.

In our next model, we added a dummy variable to indicate if teachers were in Year 2 of K-PALS implementation (Second). We hypothesized that teachers in Year 2 would have higher fidelity of implementation because of a "double dose" of the program. To obtain an estimate of the effect of the year and site of implementation on the effects of the levels of teacher support on teachers' fidelity of implementation, we introduced year and site dummy variables at Level 2 as predictors of the intercept and the slope coefficients of the dummy variables for the booster and

helper group contained in the vector G_{ij} so that the model in Equation (3) became

$$\text{avgfid}_{ij} = \beta_{0j} + G_{ij}\beta_{pj} + \beta_{2j}\text{Second} + r_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{Year} + \gamma_{02}\text{Minnesota} + \gamma_{03}\text{SouthTexas} + u_{0j}$$

$$\beta_{pj} = \gamma_{10} + \gamma_{11}\text{Year} + \gamma_{12}\text{Minnesota} + \gamma_{13}\text{SouthTexas} + u_{1j}. \qquad (4)$$

In our final model, we estimated effects of teachers' perceptions of school context on their fidelity of implementation. To the Level 1 model found in Equation (3), we added a vector C_{ij}, including Instructional Coherence; Teacher Community; Principal Leadership; the three teacher efficacy variables, Efforts and Ability, Attitudes and Habits, and School Resources; and the two K-PALS-specific scales of Instructional Coherence and Principal Leadership. Our Level 1 model now became

$$\text{avgfid}_{ij} = \beta_{0j} + G_{ij}\beta_{pj} + C_{ij}\beta_{qj} + \beta_{qj}\text{Second} + r_{ij} \qquad (5)$$

with Level 2 remaining the same as in Equation (4).

Results

To estimate the effects of K-PALS treatment conditions (i.e., the three levels of teacher support and controls), sites, and fidelity of teachers' implementation on students' reading achievement, we estimated two multilevel models (see Table 3). The first model regressed RLS on student characteristics, K-PALS treatment conditions, and school characteristics (i.e., site and Years 1 and 2). The regression coefficients for this first model are shown in column 2 of Table 3, and the standard errors appear within parentheses in column 3. The second model adds the fidelity-of-implementation measure, and results are shown in columns 4 (coefficients) and 5 (standard errors). As stated, there were three levels to each of the two models: Level 1 included the student characteristics; Level 2, teacher (or classroom) characteristics; and Level 3, school characteristics (site location and study year).

Results from the second level (teacher or classroom characteristics) were of greatest interest for

TABLE 3

Effects of K-PALS Treatment Conditions and Implementation Fidelity on Rapid Letter Sound (RLS) Reading Achievement

Variable	Model 1		Model 2	
	Coefficient	s.e.	Coefficient	s.e.
Level 1: Students (*n* = 2,959)				
Race–ethnicity				
African American	0.10	(0.90)	0.25	(0.90)
Hispanic/Latino	0.19	(1.20)	0.11	(1.18)
Asian	2.38	(1.80)	2.36	(1.79)
Other	1.00	(1.76)	0.98	(1.74)
Female	2.25	(0.43)*	2.24	(0.42)***
English language learner	–0.48	(0.99)	–0.51	(0.98)
Free/Reduced-price lunch	–1.41	(0.76)*	–1.24	(0.77)
Individualized Education Plan	–3.53	(1.26)***	–3.58	(1.26)***
Weeks between testing times	1.13	(0.11)***	1.13	(0.11)***
Level 2: Teachers (*n* = 259)				
Intercept (Control)	14.69	(3.48)***	25.43	(4.69)***
Workshop	11.10	(3.63)***	–1.37	(4.87)
Booster	18.67	(3.42)***	4.28	(5.46)
Helper	16.11	(3.42)***	1.57	(5.46)
Average fidelity			0.16	(0.04)***
Level 3: Schools (*n* = 67)				
Control (Intercept)				
Year	1.19	(1.77)	1.21	(1.76)
Minnesota	6.33	(2.01)***	6.32	(2.02)***
South Texas	2.55	(2.58)	2.63	(2.58)
Workshop				
Year	–0.75	(2.68)	–1.06	(2.47)
Minnesota	–2.87	(2.83)	–2.64	(2.68)
South Texas	–7.45	(2.43)***	–7.67	(2.31)***
Booster				
Year	–4.21	(2.21)*	–4.02	(2.11)*
Minnesota	–6.78	(2.57)***	–6.58	(2.56)**
South Texas	–7.70	(2.35)***	–7.25	(2.36)***
Helper				
Year	–1.86	(1.83)	–2.09	(1.93)
Minnesota	–10.13	(3.16)***	–9.47	(3.15)***
South Texas	–9.27	(2.40)***	–9.35	(2.55)***

Note. K-PALS = Kindergarten Peer Assisted Learning Strategies.
*$p < .10$. **$p < .05$. ***$p < .01$.

our purposes, because they show coefficients and standard errors for the effects of K-PALS treatment conditions on students' RLS scores. Findings from Model 1 in Table 3 show that as estimated by the intercept, the average control student was predicted to have an RLS gain score of 14.69 points when controlling for student and site characteristics. From the estimates of the levels of support we see that all of the groups have higher predicted average RLS gains than the control group. Also of note is that the average booster group student is predicted to have a RLS gain of 18.67 points (*ES* = 1.18) compared to the average helper group student with a gain of 16.11 points (*ES* = 1.02). This difference was tested as a General Linear Hypothesis Test of the null hypothesis that booster = helper. This test rejected the null hypothesis with $\chi^2 = 31.17$, $p < 0.0000$. All other group differences were also tested and all were similarly rejected as above.

Table 3 also displays the coefficients for Level 3, which show site differences with respect to the effects of K-PALS treatment conditions on students' RLS gains. Compared with controls in Nashville, controls in Minnesota were predicted to have a 6.33-point greater gain score on the RLS when controlling for student and treatment conditions. In addition, when examining the site estimates by workshop, booster, and helper conditions, students in South Texas had lower RLS gains than students in Nashville, ranging from 7.45 points lower in the workshop condition to 9.27 points lower in the helper condition. Minnesota students in control and workshop groups had similar gains as their counterparts in Nashville, whereas Minnesota students in booster and helper groups had inferior gain scores compared with Nashville students in booster (6.78 points lower) and helper (10.13 points lower) conditions.

Model 2 of Table 3 (see columns 3 and 4) is the same as Model 1, except we added the measure of the teachers' fidelity of implementation. With the introduction of fidelity scores to the model, estimates of the coefficients for the support group dummies are significantly reduced in magnitude and become statistically insignificant. This indicates that the effects of the K-PALS program, as differentiated by levels of support, are substantially mediated by the fidelity with which teachers implement the K-PALS program. In other words, the large effects of K-PALS treatment conditions on RLS gain scores in Model 1 may have worked through the differential fidelity associated with the varying levels of support provided to teachers as they implemented the program.

In Model 2, there continued to be site effects on RLS gain scores. South Texas children had lower RLS gains than Nashville children in workshop (7.67 points lower), booster (7.25 points), and helper (9.35 points) conditions. Compared to Nashville, Minnesota students scored 6.58 points lower in the booster condition and 9.47 points lower in the helper condition than Nashville students.

Effects of K-PALS and Setting Characteristics on Fidelity of Implementation

In addition to the effects of levels of teacher support for K-PALS, the fidelity of its implementation, and site on RLS gain scores, we examined the effects of levels of teacher support and setting characteristics on fidelity of implementation. Table 4 presents the results of our three models that estimate the effects of teacher support conditions, teachers' perceptions of school context, and site on teachers' fidelity of K-PALS implementation. In Model 1, we regressed fidelity on the three levels of teacher support (i.e., workshop, booster, and helper conditions). In Model 2, we examined effects of site and year of implementation on fidelity, controlling for levels of teacher support and whether teachers were in the 2nd year K-PALS implementation. In Model 3, we added teacher perceptions of their school context.

When considering only the K-PALS levels of teacher support, fidelity varied significantly by teachers' level of support. As expected, Model 1 shows that teachers in the helper group had highest fidelity scores, averaging 11.11 ($SD = 0.96$) points higher than the workshop group. Teachers in the booster group were predicted to have about 6 points lower fidelity than the helper group ($SD = 0.50$) but 5.37 points ($SD = 0.47$) higher than the workshop group.

When we controlled for the site of implementation and whether a teacher was implementing K-PALS for a 2nd year (Model 2), the estimates of teacher fidelity for booster and helper groups approached equal magnitude in their respective effects on fidelity scores, with the estimated effect of the booster condition increasing from 5.37 to 11.67. The estimate of the helper condition was roughly maintained, with an increase from 11.11 to 12.67. A general linear hypothesis test testing the null hypothesis that the effect of booster condition was equivalent to the effect of the helper condition on fidelity was rejected at the 0.05 level with $\chi^2 = 8.32$ and a p value < .02.

Finally, in Model 3 of Table 4, we included all variables related to the teachers' perceptions of the school context in which they work. With the exception of the estimate of the teacher efficacy variable Attitudes and Habits (the teacher's belief that students' attitudes and habits that they bring with them to school do not reduce their chances of academic success), none of the other variables was statistically significant. From this model, we would predict that a teacher who reports a point of higher agreement with this

TABLE 4

Effects of K-PALS Treatment Conditions and Setting Characteristics on Fidelity of Implementation

Variable	Model 1		Model 2		Model 3	
	Coefficient	s.e.	Coefficient	s.e.	Coefficient	s.e.
Level 1: Teachers (*n* = 197)						
Workshop (Intercept)	81.38	(1.24)***	79.64	(4.38)***	79.78	(4.23)***
Booster	5.37	(1.58)***	11.67	(4.54)**	11.22	(4.25)***
Helper	11.11	(1.82)***	12.61	(5.62)**	11.43	(5.86)*
Years teaching					−0.01	(0.09)
Efforts and ability					−0.26	(1.10)
Attitudes and habits					2.03	(1.07)*
School resources					−0.06	(1.35)
Second			−2.18	(2.01)	−1.70	(2.06)
Instructional coherence					−2.98	(2.92)
Teacher community					1.49	(1.94)
Principal leadership					2.07	(2.34)
PALS instructional coherence					0.76	(1.79)
PALS principal leadership					−1.61	(1.30)
Level 2: Schools (*n* = 61)						
Workshop (Intercept)						
Year			1.53	(3.04)	1.36	(2.98)
Minnesota			−1.54	(3.49)	−2.14	(3.46)
South Texas			1.53	(3.00)	2.19	(3.24)
Booster						
Year			−3.40	(3.97)	−2.63	(3.81)
Minnesota			−1.67	(3.43)	−1.75	(3.12)
South Texas			−4.31	(2.68)	−5.38	(3.00)
Helper						
Year			−0.67	(3.18)	0.65	(3.44)
Minnesota			−0.86	(4.88)	−3.01	(4.89)
South Texas			−0.97	(3.57)	−2.14	(4.12)

Note. K-PALS = Kindergarten Peer Assisted Learning Strategies.
*$p < .10$. **$p < .05$. ***$p < .01$.

statement than the mean of teacher responses in the sample would have an estimated average fidelity 2.03 points higher, controlling for level of support and other perceptions of school context. The magnitude of this estimate is quite modest, with an estimated effect size of 0.18 on average fidelity. There were no other substantive differences when comparing Model 3 and the more parsimonious Model 2.

Discussion

We have described a portion of our results from the first 2 years of a randomized control trial to explore how best to scale up K-PALS, a so-called best-evidence program to accelerate reading development among young children. This 2-year experimental study analysis involved 2,959 students and 259 teachers in 67 urban, suburban, and rural schools in Nashville, Minnesota, and six of America's poorest school districts in South Texas. Our findings show generally that the level of teacher support for K-PALS implementation is important for early reading achievement gains. Estimated effects of level of teacher support are robust—even with significant differences across the three study sites. Evidence suggests that much of the gains achieved are mediated by the fidelity with which teachers implemented the K-PALS program. We wish to emphasize in this regard that our one early reading measure was RLS. Although many

reading researchers and practitioners would agree on the theoretical and practical relevance of this measure to early reading development—and despite that letter sounds are an important focus of K-PALS—we recognize that this study's importance would be enhanced had we examined relations between level of teacher support, fidelity of implementation, and children's performance by using additional measures of reading at the word level and in connected text.

Our findings also suggest that program characteristics matter. As indicated, the K-PALS program is highly structured, providing specific plans for teachers. K-PALS teachers can access manuals and materials to guide them through the various reading activities. Specified amounts of time each week are part of the intervention. As we hypothesized, these program characteristics seem related to fidelity, as evidenced by the high levels of fidelity across all three treatment conditions. This raises an obvious question, namely, whether interventions of a contrasting nature, and with different objectives, may be brought to scale in the same way that we scaled K-PALS. Early science instruction, or strategies to strengthen reading comprehension, as but two examples, may require alternate approaches to scaling up.

As important as program characteristics may be, additional considerations are likely to affect fidelity of treatment implementation and student achievement. Our randomized control trial was designed so that we could explore experimentally whether level of teacher support also influenced the fidelity with which K-PALS was implemented and affected students' early reading performance. Our most basic (i.e., minimal) level of support was a workshop condition during which teachers spent a day in training, sharing ideas, and familiarizing themselves with the program and its materials. A more intensive support condition was involving teachers in two booster sessions with K-PALS staff, in addition to their workshop participation, during which questions were answered, problems were discussed, and ideas were shared. Finally, teachers in what we initially conceived as our most intensive helper condition received the workshop and booster sessions and a research assistant (i.e., helper) who visited weekly to assist with student training and K-PALS implementation.

Whereas our analyses indicated that all three levels of teacher support were related to students' early reading gains compared to controls, the respective effects of the three levels of support were not entirely expected. We hypothesized that as level increased from workshop to booster to helper, the effects of each on children's performance would also increase. In fact, we found that students in the helper condition did not outperform students in the booster condition. In part, this may be because of the inconsistent quality of assistance provided by graduate students to teachers. Such help and mentoring requires a highly trained K-PALS professional, and graduate students new to the K-PALS program may not be in the best position to provide the type of assistance to help teachers outperform the teachers in the booster condition, which was run by experienced K-PALS staff. Moreover, there was anecdotal evidence from the Minnesota site that at least several classroom teachers resented the presence of the relatively inexperienced K-PALS helpers.

Thus, the disappointing news is that the intentional structuring of supportive relationships with teachers did not result in the expected outcome, although the helper condition did have positive effects compared with control and workshop conditions. The good news is that the helper condition, although important in principal in providing high-quality professional development (see Desimone, Porter, Garet, Suk Yoon, & Birman, 2002), is expensive. So, the finding that the booster condition facilitates meaningful effects on early reading achievement suggests a cost-effective means of intervention support.

In addition, our analyses revealed that when teachers implemented K-PALS with fidelity, students gain in reading. In fact, once fidelity was taken into account, the effects of the three levels of teacher support appeared to be significantly mediated as hypothesized. In addition, our unadjusted model indicated that teachers in the booster condition had higher fidelity scores than teachers who participated only in the workshop. Moreover, teachers in the helper condition had higher levels of fidelity than teachers in booster or workshop conditions.

However, this leads to the other aspect of the theoretical framework we offered earlier—the social context, or setting. As a randomized study

at scale, our investigation extended K-PALS implementation to Minneapolis, St. Paul, and Bloomington (Minnesota) and six contiguous districts in South Texas. In addition, we examined several teacher characteristics that might mediate the effects of fidelity of implementation (e.g., experience, sense of efficacy, and perceptions about classroom and school climate) and classroom features. When analyzing the effects of these setting characteristics, we found no individual measures of site to be significantly related to fidelity, but together, the site measures added significantly to model fit. In addition, site seemed to moderate effects of booster and helper conditions on fidelity scores compared to the workshop condition. That is, by controlling for site, booster and helper conditions became more similar compared to the unadjusted effects.

For implementation of interventions at scale, variation in the fidelity with which treatments are conducted presents a significant challenge (Glennan et al., 2004). A common approach is to recruit high-quality staff to provide implementation support (e.g., Slavin & Madden, 2007), although degree of implementation may still fluctuate because of the varying quality of the ongoing technical assistance, even with highly structured programs (Berends et al., 2002). This is a continuing challenge within our study, something that we will address in the coming years.

Although theory led us to hypothesize that teacher experience, self-efficacy, and perceptions of school and classroom climate would be related to fidelity scores, we found few such effects. In part, this supports the group equivalency within our study design, whereby we randomly assigned teachers to study groups, suggesting that teacher characteristics were randomly distributed across treatment and control conditions. By and large, our results reveal this is the case. However, we did find that teachers with higher fidelity scores tended to believe that their students' attitudes and habits did not reduce their chances of academic success.

Overall, our results show that within the context of a randomized field trial, it is easier to test effects of various, intentionally modified program characteristics and setting characteristics across sites than it is to test variation in teacher attitudes. In future research, we will examine whether teachers

sustain implementation of K-PALS, and in this regard, we will continue to address the theoretical framework that emphasizes teacher perceptions about their school contexts as well as program characteristics.

In conclusion, very few studies of evidence-based instructional programs document high-quality measures of fidelity, especially within a randomized design to study program effects at scale. In this study, we had an opportunity to examine effects of a highly structured, skills-oriented program, while accounting for how they might vary by level of teacher support and fidelity of implementation, at various sites. As we pursue our research agenda, we hope to contribute to thinking about how various program characteristics, support features, and site factors influence the structuring of relationships between teachers and students on specific curricular content and how they affect early reading achievement. As Elmore (2004) observes about the problem of scaling-up effective educational practices,

> Recent efforts to improve instructional practice at scale focus more on the number of adopters and the structural characteristics of reform than they do on fundamental changes in the instructional core—the relationships between teachers and students and the organizational practices that support those relationships. The difficulty of making changes at this level is the principal constraint on the large-scale adoption of promising new practices. (p. 8)

Our results suggest the importance of several factors in bringing evidenced-based instructional practices to scale, including the programmatic features of the intervention, the levels of on-site teacher support, and site. Yet given the particular nature of our educational intervention, and that we used only one measure of early reading improvement, more research and development is necessary to understand how to strengthen teacher support and fidelity of treatment implementation as we endeavor to bring evidence-based instructional practices to scale.

References

Barton-Arwood, S. M., Falk, K. B., & Wehby, J. H. (2005). Reading instruction for elementary-age students with emotional and behavioral disorders:

Academic and behavioral outcomes. *Exceptional Children*, *72*(1), 7–27.

Berends, M. (2004). In the wake of *A Nation at Risk:* New American School's private sector school reform initiative. *Peabody Journal of Education*, 79(1), 130–163.

Berends, M., Bodilly, S., & Kirby, S. N. (2002). Facing the challenges of whole-school reform: New American Schools after a decade. Santa Monica, CA: RAND.

Berman, P., & McLaughlin, M. (1976). Implementation of educational innovation. *Educational Forum*, *40*, 345–370.

Berman, P., & McLaughlin, M. (1978). *Federal programs supporting educational change: Vol. 8. Implementing and sustaining innovations*. Santa Monica, CA: RAND.

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, *77*(3), 211–244.

Desimone, L., Porter, A. C., Garet, M., Suk Yoon, K., & Birman, B. (2002). Effects of professional development on teachers' instruction: Results from a three-year study. *Educational Evaluation and Policy Analysis*, *24*(2), 81–112.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, *18*(2), 237–256.

Elmore, R. (2004). *School reform from the inside out: Policy, practice, and performance*. Boston: Harvard University Press.

Elmore, R. (1996). Getting to scale with good educational practice. *Harvard Educational Review*, *66*(1), 1–26.

Fuchs, D., & Fuchs, L. S. (1998). Researchers and teachers working together to adapt instruction for diverse learners. *Learning Disabilities Research and Practice*, *13*, 126–137.

Fuchs, D., Fuchs, L. S., Al Otaiba, S., Thompson, A., Yen, L., McMaster, K. N., et al. (2001). K-PALS: Helping kindergartners with reading readiness. Teachers and researchers in partnerships. *Teaching Exceptional Children*, *33*(4), 76–80.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-Assisted Learning Strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*, 174–206.

Fuchs, D., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N. J., et al. (2001). Is reading important in reading readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology*, *93*(2), 251–267.

Fuchs, D., Fuchs, L. S., Thompson, A., Yen, L., Al Otaiba, S., Nyman, K., et al. (2001). Peer-assisted learning strategies in reading: Extensions for kindergarten, first grade, and high school. *Remedial and Special Education*, *22*, 15–21.

Fullan, M. (2001). *The new meaning of educational change* (3rd ed.). New York: Teachers College Press.

Fullan, M., & Pomfret, A. (1977). Research on curriculum and instruction implementation. *Review of Educational Research*, *47*, 335–397.

Gersten, R., Chard, D., & Baker, S. . (2000). Factors that enhance sustained use of research-based instructional practices: A historical perspective on relevant research. *Journal of Learning Disabilities*, *33*(5), 444–457.

Glennan, T. K., Bodilly, S. J., Galegher, J. R., & Kerr, K. A. (Eds.). (2004). *Expanding the reach of education reforms: Perspectives form leaders in the scale-up of educational interventions*. Santa Monica, CA: RAND.

Gottfredson, G. (1984). A theory-ridden approach to program evaluation: A method for stimulating researcher-implementer collaboration. *American Psychologist*, *39*, 1101–1112.

Huberman, M. (1995). Professional careers and professional development: Some intersections. In T. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 193–224). New York: Teachers College Press.

Levy, B. A., & Lysunchuk, L. (1997). Beginning word recognition: Benefits of training by segmentation and whole word methods. *Scientific Studies of Reading*, *1*, 359–387.

Louis, K. S., & Marks, H. M. (1998). Does professional community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education*, *106*(4), 532–575.

Moncher, F., & Prinz, R. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*(3), 247–266.

Newmann, F. M., Smith, B. A., Allensworth, E., & Bryk, A. S. (2001). Instructional program coherence: What it is and why it should guide school improvement policy. *Educational Evaluation and Policy Analysis*, *23*(4), 297–321.

Rohrbach, L., Graham, J., &Hansen, W. (1993). Diffusion of a school-based substance abuse prevention program: Predictors of program implementation. *Preventive Medicine*, *22*(2), 237–260.

Ruiz-Primo, M. (2005). *A multi-method and multi-source approach for studying fidelity of implementation*. Paper presented at the annual meeting of the American Education Research Association, Montreal, Canada.

Schneider, B., & McDonald, S. K. (2007). *Scale-up in education: Issues in practice* (Vol. 2). New York: Rowman and Littlefield.

Slavin, R. E., & Madden, N. A. (2007). Scaling up Success for All: The first sixteen years. In B. Schneider & S. McDonald (Eds.), *Scale-up in education: Issues in practice* (Vol. 2, pp. 201–227). New York: Rowman and Littlefield.

Smylie, M. A. (1988). The enhancement function of staff development: Organizational and psychological antecedents to individual teacher change. *American Educational Research Journal*, *25*, 1–30.

Vaughn, S., Klingner, J., & Hughes, M. (2000). Sustainability of research-based practices. *Exceptional Children*, *66*(2), 163–171.

Weatherly, R., &.Lipsky, M. (1977). Street level bureaucrats and institutional innovation: Implementing special education reform. *Harvard Education Review*, *47*(2), 171–197.

Authors

MARC L. STEIN is currently an advanced doctoral student in the Department of Leadership, Policy and Organizations and is a fellow with the Experimental Education Research Training Program (ExpERT), which is supported by the Institute for Education Sciences (IES), at Vanderbilt University, Peabody College, Box 459, GPC, 230 Appleton Place, Nashville, TN 37203; marc.1.stein@vanderbilt.edu. His current research focuses on social contexts and their relationship to differential seasonal learning, education policy and teaching practice.

MARK BERENDS is a professor of sociology at Notre Dame and director of the Center for Research on Educational Opportunity and director of the National Center on School Choice at Vanderbilt University, Peabody College (http://www.vanderbilt .edu/schoolchoice/). He also serves as the vice president of the American Educational Research Association's Division L, Policy and Politics in Education. His areas of expertise are the sociology of education, research methods, school effects on student achievement, and educational equity. Mailing address: Department of Sociology, University of Notre Dame, 810 Flanner, Notre Dame, Indiana 46556.

DOUGLAS FUCHS is a Nicholas Hobbs Professor of Special Education and Human Development at Peabody College of Vanderbilt University, where he is also codirector of the Vanderbilt–Kennedy Center Reading Clinic. His research interests and policy work address the academic needs of elementary-grade children who are at risk because of low income or disability. Mailing address: Special Education Department, Peabody College Box 228, 110 Magnolia Circle, 418C MRL, Nashville, TN 37203.

KRISTEN L. McMASTER is an assistant professor of special education in the Department of Educational Psychology, University of Minnesota, 56 East River Rd., 250 Education Sciences Bldg., Minneapolis, MN 55455. Her research interests include creating conditions for successful response to academic interventions of students at risk and students with disabilities. Specific research focuses on (a) promoting teachers' use of data-based decision making and evidence-based instruction and (b) developing individualized interventions for students for whom generally effective instruction is not sufficient.

LAURA SÁENZ is an assistant professor of special education at University of Texas Pan American, College of Education, Department of Educational Psychology, 1201 W. University Drive, Edinburg, TX 78539. She specializes in literacy instruction for students with learning disabilities with special emphasis on English language learners. She is coprincipal investigator and coordinator of the IES Scaling Up Peer Assisted Learning Strategies (PALS) grant for the South Texas site.

LOULEE YEN has been a project coordinator for Dr. Doug Fuchs for more than 10 years. She has co-ordinated large-scale Reading PALS studies of kindergarten through fifth-grade PALS. She currently coordinates the Nashville site of the IES Scaling Up grant and coordination between the other research sites in Minneapolis and South Texas. Mailing address: Vanderbilt University, Special Education Department, Peabody College Box 228, 110 Magnolia Circle, 418C MRL, Nashville, TN 37203.

LYNN S. FUCHS is a Nicholas Hobbs Professor of Special Education and Human Development at Vanderbilt University, 228 Peabody, Nashville, TN 37203. Her research concerns classroom formative assessment, reading and mathematics instruction, and learning disabilities.

DONALD L. COMPTON is an associate professor of special education and a John F. Kennedy Center Investigator at Peabody College, Vanderbilt University, Special Education Department, Peabody College, Box 228, 110 Magnolia Circle, 418C MRL, Nashville, TN 37203. His research involves modeling individual differences in the development of children's reading skills and the exploration of alternative methods for identifying children with reading disabilities.