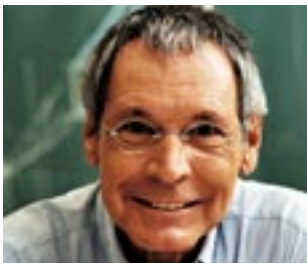
ISSUES&ANSWERS REL 2009–No. 082

REL NORTHWEST
Regional Educational Laboratory
At Education Northwest

New and experienced teachers in a school reform initiative: the example of Reading First



NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

ISSUES & ANSWERS REL 2009–No. 082

REL NORTHWEST
Regional Educational Laboratory
At Education Northwest

New and experienced teachers in a school reform initiative: the example of Reading First

November 2009

Prepared by

Kari Nelsestuen
Education Northwest

Caitlin Scott
Education Northwest

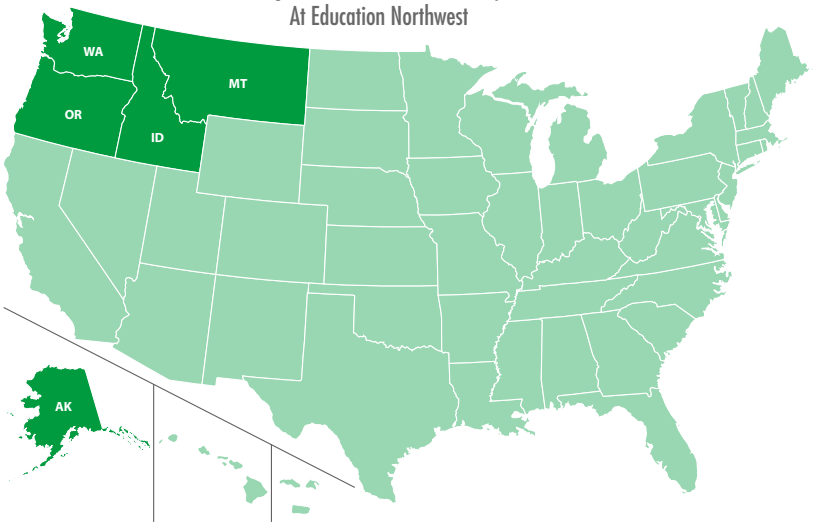
Makoto Hanita
Education Northwest

LeAnne Robinson
Western Washington University

Tracy Coskie
Western Washington University

NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL NORTHWEST
Regional Educational Laboratory
At Education Northwest



Issues & Answers is an ongoing series of reports from short-term Fast Response Projects conducted by the regional educational laboratories on current education issues of importance at local, state, and regional levels. Fast Response Project topics change to reflect new issues, as identified through lab outreach and requests for assistance from policymakers and educators at state and local levels and from communities, businesses, parents, families, and youth. All Issues & Answers reports meet Institute of Education Sciences standards for scientifically valid research.

November 2009

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-06-CO-0016 by Regional Educational Laboratory (REL) Northwest administered by Education Northwest. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Nelsestuen, K., Scott, C., Hanita, M., Robinson, L., and Coskie, T. (2009). *New and experienced teachers in a school reform initiative: the example of Reading First* (Issues & Answers Report, REL 2009–No. 082). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. Retrieved from www.ies.ed.gov/ncee/edlabs.

This report is available on the regional educational laboratory web site at www.ies.ed.gov/ncee/edlabs.

Summary

REL 2009–No. 082

New and experienced teachers in a school reform initiative: the example of Reading First

This study compares the experiences and perceptions of new and experienced teachers in four key areas of school reform. Data from 235 Reading First schools in six western states revealed differences in teachers' experiences with reading coaches (but not in perceptions of support from coaches) and in teachers' confidence in using data to guide instruction, but no differences in their perceptions of collaborative grade-level meetings or overall support for Reading First.

This study examines the experiences and perceptions of new and experienced teachers in Reading First, a federal school reform initiative implemented in more than 5,880 high-poverty, low-performing schools across the country. The initiative aims to help schools improve reading among K–3 students with the goal of having all students reading at grade level by the end of grade 3. Reading First is the largest federal reading initiative in history, and its influence has spread to many other schools in the region and across the country (Deussen, Nelsestuen, and Scott 2008; Scott 2006).

The study considers four areas of the multifaceted Reading First reform model: instructional coaches, teacher collaboration, teacher use of student assessment data, and support for reform. The first three elements are required in all Reading First schools and are commonly found in other reform models as well (Comprehensive School Reform Quality Center 2005). While there is some literature about these reform elements, little attention has been given to new teachers' experiences with reform. The issues are particularly relevant in Reading First schools, which tend to have a high proportion of new teachers (in their first four years of teaching). Some 34 percent of teachers are new in Reading First schools in the six western states (Alaska, Arizona, Idaho, Montana, Washington, and Wyoming) included in this study compared with 22 percent across all elementary schools in the same states (U.S. Department of Education 2006).

Four research questions drove this study:

- How do new teachers' experiences with their reading coach differ from those of their more experienced peers?
- How do new teachers' experiences with teacher collaboration differ from those of their more experienced peers?
- How do new teachers' perceptions of their data skills differ from those of their more experienced peers?

- How does new teachers' support for the reform initiative differ from that of their more experienced peers?

To answer these questions, the study examined previously collected data from all 235 Reading First schools in the six western states. The data came from 2008 statewide evaluations and included teacher surveys, teacher interviews, and coach interviews. Surveys were collected from 3,094 (91 percent) of the K–3 teachers in these schools. The study analyzed a subset of 19 survey items using hierarchical linear modeling to test for differences between the responses of new and experienced teachers. Interview data came from a subset of 169 teachers and 85 reading coaches from 85 (36 percent) of the 235 schools.

The first question focused on instructional coaching, a practice gaining in popularity across the country (Russo 2004). Reading coaches, as they are called in Reading First programs, are school staff members responsible for providing ongoing professional development to teachers through modeling, observing, and providing feedback to teachers. The study found three differences in new teachers' experiences with their reading coach. First, the probability of teachers reporting feedback occurring once a month or more was higher for new teachers than for experienced teachers, and the response difference was significant ($t[5260] = 2.53$, $p < 0.05$). Second, interviewed coaches reported that new teachers needed more help with classroom management, core program use, and modeling of teaching practices. Third, a higher percentage of interviewed new teachers believed that coaching assistance had changed their instruction. Despite these differences, however, there were no differences between new and experienced teachers' overall perceptions of the effectiveness of support from their reading coach, based on survey data ($t[178] = 1.12$, $p > .05$).

The second question focused on teacher collaboration, an increasingly common practice in the school reform movement (Johnson, Berg, and Donaldson 2005). The mechanism for collaboration in most Reading First schools is the grade-level meeting at which teachers discuss curriculum, instruction, and the needs of individual students. No differences were found between new and experienced teachers' views of the usefulness of collaborative grade-level meetings; both surveyed groups agreed that the meetings were a good use of their time. Teacher interview data confirmed this finding.

Teachers' perceptions of their data skills was the focus of the third question. Regular use of student assessment data is now an almost universal expectation in school reform efforts (Scott 2007). In Reading First, teachers are expected to regularly use the results of reading assessments to make instructional decisions, such as on grouping students and differentiating lessons. Survey results indicate that new teachers' confidence in their ability to use data for tasks such as grouping students and understanding schoolwide trends was significantly lower than that of experienced teachers ($t[2200] = -8.55$, $p < 0.01$).

The final question examined teachers' support for the Reading First reforms. Support—or its absence—can influence the likelihood of sustaining or scaling up reform (Steiner 2000; Taylor 2005). Both survey and interview data indicated that new teachers' support for

Reading First was similar to that of experienced teachers.

Since data for this study came from Reading First schools in western states, findings may not be generalizable to other regions or other reform initiatives. This study is also limited by its descriptive nature; it says nothing about whether differences between new and experienced teachers had any ultimate impact on teacher instruction or student learning. A descriptive study also cannot establish whether the differences between new and experienced teachers are due to differences in years of teaching experience, since other factors not accounted for in the study could contribute to how teachers experience and react to school reform.

November 2009

TABLE OF CONTENTS

Why this study? 1
What is Reading First? 1
Need for the study and study questions 2

Study findings 3
Overview of findings 3
How do teachers' experiences with reading coaches differ? 5
How do teachers' experiences with collaboration differ? 7
How do teachers' perceptions of their data skills differ? 8
How does teachers' support for the reform initiative differ? 9

Study limitations 10

Appendix A What we know from the literature 11

Appendix B Data sources and study methods 14

Notes 26

References 27

Boxes

1 Terminology used in this study 2

2 Study methods 4

Tables

1 Summary of findings 5

2 Frequency of coach feedback reported by new and experienced teachers, spring 2008 6

3 New and experienced teachers' perceptions of support from their reading coach, spring 2008 7

4 New and experienced teachers' views of the usefulness of collaborative meetings, spring 2008 8

5 New and experienced teachers' confidence in their data skills, spring 2005 9

6 New and experienced teachers' support for Reading First, spring 2008 9

B1 Teacher survey items, spring 2008 15

B2 Scales from survey items by research area, spring 2008 16

B3 Teacher survey response rates, by state, spring 2008 17

B4 Number of teacher survey respondents at two levels of teaching experience, by state, spring 2008 17

B5 Interview items for coaches and teachers, by research area, spring 2008 18

B6 Interview sample size and selection method, by state, spring 2008 18

B7 Total number of interview respondents for each question, by level of experience, spring 2008 19

B8 Characteristics of participating schools, 2003/04 19

B9 Characteristics of surveyed and interviewed teachers, spring 2008 21

B10 Results of random analysis of variance model for calculating the intraclass correlation coefficient 21

B11 Interview data analysis code structure 23

This study compares the experiences and perceptions of new and experienced teachers in four key areas of school reform. Data from 235 Reading First schools in six western states revealed differences in teachers' experiences with reading coaches (but not in perceptions of support from coaches) and in teachers' confidence in using data to guide instruction, but no differences in their perceptions of collaborative grade-level meetings or overall support for Reading First.

WHY THIS STUDY?

This study examines the experiences and perceptions of new and experienced teachers in Reading First, a federal school reform initiative for improving reading outcomes for K–3 students in high-poverty, low-performing schools. Since 2003, Reading First has given more than $6 billion to states, which gave grants to more than 5,800 schools to provide professional development to teachers, purchase scientifically based reading materials, and promote the use of reading assessment data to make informed instructional decisions (Gamse et al. 2008). The ultimate goal of Reading First is to ensure that all students read at grade level by the end of grade 3.

What is Reading First?

Reading First is similar to many other school reform models (for definition of this and other key terms, see box 1). Reading First schools are required to hire a reading coach to work with teachers, have teachers collaborate through grade-level meetings, and use data from regular student assessments to make instructional decisions. Other popular reform models implemented in schools across the country also contain these elements.

Instructional coaching, for example, is practiced increasingly across the country (Russo 2004). Coaches are often skilled teachers who step out of their classrooms to help other teachers become more thoughtful and effective in their instruction. Coaches work with teachers in the classroom, observing, modeling, providing feedback, and planning lessons, according to each teacher's needs. The intention is to address professional development for high-quality teachers through coaching that is job-embedded, ongoing, directly related to the challenges teachers face, and provided by people familiar with the context of the teachers' work. In a 2005 study of 22 comprehensive reform models by the Comprehensive School Reform Quality Center, coaching was a noted feature of at least seven models implemented in more than 4,000 schools nationally, such as America's Choice,

BOX 1

Terminology used in this study

Data skills. Skills in using student assessment data to inform instructional decisions, such as how to group students and what to teach.

Experienced teacher. Teacher with five or more years of experience.

New teacher. Teacher with one to four years of experience.

Reading coach. Staff member responsible for providing professional development and support to teachers to improve their reading instruction.

School reform model. Specific structures and practices packaged together for adoption by schools seeking change.

Teacher collaboration. Formal collaboration through regular grade-level meetings—required in Reading First schools.

Modern Red School House, and Coalition of Essential Schools.[1]

Teacher collaboration, a second required component of Reading First, occurs through regular grade-level meetings among teachers to plan lessons, examine student data, and address other issues of teaching and learning. The school reform movement has been credited with increasing the emphasis on teacher collaboration (Johnson, Berg, and Donaldson 2005). The Comprehensive School Reform Quality Center (2005) found that collaboration was a required component of nine school reform models implemented in more than 2,700 schools across the country, including Atlas Communities, First Steps, and Modern Red School House.

Reading First schools must also assess students regularly and use the resulting data to plan instruction, student grouping, and professional development. Coaches, principals, and teachers are all expected to understand and use results from multiple reading assessments. The regular use of data is also part of 21 of 22 other reform models implemented in almost 9,000 schools across the country (Comprehensive School Reform Quality Center 2005).

While Reading First is similar to other reform models, it stands out for its unprecedented reach—more than 5,800 schools participate, and components of Reading First have spread to other schools in the same districts (Scott 2006). In addition, Reading First promotes three-tiered instruction, a model increasingly disseminated in response to intervention initiatives.[2] For these reasons, lessons learned in Reading First schools may be relevant for other schools, especially those seeking to adopt schoolwide reforms with similar program elements.

Need for the study and study questions

This study is particularly relevant to policymakers, administrators, and educators in the six western states—Alaska, Arizona, Idaho, Montana, Washington, and Wyoming—that are the focus of this study. Reading First has been implemented in more than 230 schools in these states, and elements of Reading First such as coaching and use of data have spread to other schools in the states (Deussen, Nelsestuen, and Scott 2008).

The study compares the experiences and perceptions of new teachers (one to four years of teaching experience) with those of experienced teachers (five or more years of teaching experience) at Reading First schools. Reading First schools, like other high-poverty schools, have disproportionately high percentages of new teachers.[3] Across the six states from which this study drew its data, 34 percent of teachers in Reading First schools are new teachers compared with 22 percent in all elementary schools in the same states (U.S. Department of Education 2006).

This disproportionately high percentage of new teachers in Reading First schools is not surprising; other research has documented that high-poverty schools tend to have a less experienced teaching

force than do more affluent schools (Clotfelter et al. 2007; Lankford, Loeb, and Wycoff 2002). In this sense, the experience of new teachers is particularly relevant to high-poverty schools—the kind of schools that have received Reading First grants.

Little has been published on new teachers' experiences and perceptions of school reform models (see appendix A for a review of the literature). For example, only one study was found on how coaches work with new teachers (Marsh et al. 2008), and only a few studies describe or measure new teachers' use of data in making instructional decisions. State evaluations have found support for the reform among teachers (Autio, Roccograndi, et al. 2008; Autio, Scott, et al. 2008; Nelsestuen, Scott, and Burke 2008), but studies have not explored whether support varies by teachers' experience, as Hargreaves (2005) suggests that it might. Because of the lack of research, it is unclear whether reforms need to be differentiated for new and experienced teachers.

This study examines the experiences and perceptions of new teachers in Reading First and compares them with those of experienced teachers. The focus is on topics that are both central to Reading First and common to other school reform models: working with a reading coach, collaborating with other teachers, and using assessment data. The study also compares new and experienced teachers in overall support for the reforms under Reading First.

Four research questions drove this study:

- How do new teachers' experiences with their reading coach differ from those of their more experienced peers?
- How do new teachers' experiences with teacher collaboration differ from those of their more experienced peers?
- How do new teachers' perceptions of their data skills differ from those of their more experienced peers?
- How does new teachers' support for the reform initiative differ from that of their more experienced peers?

Differences in experiences and perceptions were found between new and experienced teachers in coaching and data skills but not in collaboration and support for reform

Data for the study, collected in spring 2008, included teacher surveys and teacher and coach interviews from Reading First schools in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming (see box 2 and appendix B for details on data collection and study methods).

STUDY FINDINGS

This section presents results of the study of the experiences and perceptions of new and experienced teachers in four areas central to Reading First and common to other school reform models, combining findings from survey and interview data.

Overview of findings

The changes that school reform models bring to a school can be experienced or perceived differently by different member of the school staff. In the two study areas of coaching and data skills, differences in experiences and perceptions were found between new and experienced teachers (table 1). In the other two study areas of collaboration and support for reform, there was little difference in the perceptions of new and experienced teachers.

The study found three differences in coaching between new and experienced teachers. First, the probability of teachers reporting feedback occurring once a month or more was higher for new teachers, a finding that corroborates another recent study that reported that coaches emphasized working with new teachers (Marsh et al. 2008). Second, interviewed coaches reported that new teachers needed more help with classroom

BOX 2

Study methods

The study used a concurrent mixed methods design (Creswell 2003) to analyze survey and interview data concurrently. Data from teacher surveys in the six study states and from teacher and coach interviews were collected in spring 2008. Using similar survey and interview instruments in each state made it possible to combine datasets for analysis.

Data sources. K–3 teachers in 235 fully funded Reading First schools in the six states were surveyed as part of statewide evaluations of Reading First. Of the surveys' more than 100 close-ended items, 19 were relevant to this study. Two items were used individually, while factor analysis was used to create three scales from the remaining 17 items (see table B2 in appendix B for details). Coach support was measured through an individual item about the frequency of coach observations and through a scale developed by averaging four survey items about coach support (see table B1 in appendix B for survey items used in the study). Data confidence was measured with a scale developed from averaging five survey items. The usefulness of collaboration was measured with a single survey item. Support for reform was measured using a scale developed by averaging eight survey items. All items except the frequency of coach observations used the same five-point scale.

Interviews were conducted with reading coaches and teachers at a randomly selected subset of 85 schools (36 percent of the total; see table B6 in appendix B). Each school had one Reading First coach; all 85 were interviewed. The teachers interviewed at each school were selected to include an even distribution of new and experienced teachers, yielding interviews with 84 new teachers and 85 experienced teachers.

Sample. Teacher surveys were conducted at 235 schools—all fully funded Reading First schools—in the six states during 2007/08 (table B8 in appendix B shows the characteristics of participating schools, including size, years in Reading First, and student demographics).

Data analyses. Descriptive results were calculated for all the survey items/scales of interest for the teacher survey data. Multiple imputation was used to address missing data. Intraclass correlation coefficient analysis was used to determine whether the analysis method needed to address the nested nature of the data (teachers are nested within schools, so school characteristics could lead to similarity among teachers' responses; see table B10 in appendix B). Hierarchical linear modeling was selected for analyzing the three scales (coach support, data confidence, and support for reform) and one item (usefulness of collaborative meetings) that were measured with a five-point Likert scale response; four models were created, with the predictor of interest being whether teachers were new or experienced. The outcome variables were the three scales and the one individual item (described in table B1 in appendix B). For the fifth item (frequency of coach support), hierarchical generalized linear modeling was used because the response scale was an ordinal variable with uneven spacing between points on the scale. Predicted probabilities were calculated for new and experienced teachers reporting "once a month or more" in response to this item.

Interview data from coaches and teachers were analyzed using an inductive coding process—researchers examined the data and developed codes inductively to categorize responses (see table B11 in appendix B for codes). Interrater reliability for consistency of researcher coding was 98 percent. Once data were coded, common patterns of responses were examined.

management, core program use, and modeling of teaching practices. And third, a higher percentage of new teachers believed that coaching assistance had changed their instruction. There were no differences in survey data between new and experienced teachers' overall perceptions of the effectiveness of support from their reading coach.

No differences were found between new and experienced teachers' views of the usefulness of collaborative grade-level meetings; on average, both surveyed groups agreed that the meetings were a good use of their time. Other studies have also found that teachers, regardless of experience, value collaboration (Imbimbo 2004; National Education Association 2003).

TABLE 1
Summary of findings

Area of reform implementation	Difference between new and experienced teachers?	Data source
Coaching		
Frequency of feedback to teachers from their coach	Yes. Probability of teachers reporting feedback occurring once a month or more was higher for new teachers.	Survey
Types of support new teachers need from their reading coach	Yes. Coaches noted three areas where new teachers needed more support than experienced teachers did.	Interview
Teachers' perceptions of the impact of the reading coach on their instruction	Yes. A higher percentage of new than of experienced teachers said their coach had changed their instruction.	Interview
Teachers' perceptions of the support they received from their reading coach	No.	Survey
Collaboration		
Teachers' views of the usefulness of collaborative meetings	No.	Survey and interview
Data skills		
Teachers' confidence in their skills to use student data	Yes. Confidence in data skills was lower for new teachers than for experienced teachers.	Survey
Support for reform		
Support for the school reform	No.	Survey and interview

Source: Authors' analysis of the 2008 teacher surveys and teacher and coach interviews administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

Surveyed new teachers' confidence in their data skills was lower than that of their more experienced peers. Other studies have concluded that new teachers need more professional development in data use (Datnow, Park, and Wohlstetter 2007; Marsh, Pane, and Hamilton 2006).

Finally, both survey and interview data indicate that new teachers' support for Reading First was similar to that of experienced teachers. Another recent study reached different conclusions: Hargreaves (2005) found that new teachers were more receptive to education reforms than experienced teachers were, although the study looked at reform in general rather than at specific reform models such as Reading First.

How do teachers' experiences with reading coaches differ?

The study examined four areas related to teachers' experience with reading coaches: frequency of feedback to teachers (survey data), coaches' perceptions of differences in coaching needs of new and experienced teachers (coach interview data), teachers' perceptions of the impact of coaching on instruction (teacher interview data), and teachers' perceptions of the support they received from their reading coach (teacher survey data).

Frequency of feedback to teachers. Providing regular feedback to teachers on their instruction is a key role of Reading First coaches. The states in this study expect coaches to attend multiple reading classrooms daily, eventually serving all K–3 teachers in their school. The maximum ratio of coaches to teachers in the study schools was 1:20, implying that coaches would provide feedback to each teacher at least monthly. The frequency of feedback was measured through a single teacher survey item asking: "This year, how often did the reading coach provide you with feedback on your

Thirty three percent of interviewed coaches reported that new teachers had a greater need for help with classroom management, including skills such as establishing routines and procedures for students and managing student behavior

instruction?" To reflect the expectation of at least monthly feedback from coaches, the six response options in the survey were collapsed into two: teachers reporting that they received feedback once a month or more, and teachers reporting that they received feedback less than once a month.[4]

Feedback frequency was analyzed using a hierarchical generalized linear model (see appendix B for a description). The probability of teachers reporting feedback occurring once a month or more was 69 percent for new teachers and 64 percent for experienced teachers (table 2). The difference was statistically significant.[5]

Coaches' perceptions of differences in coaching needs of new and experienced teachers. Coaches were asked: "Do new teachers have different needs than veteran teachers? How?" Of the 85 coach responses, 83 (98 percent) were used in the analysis. The two responses that were not included came from a coach who said the school had no new teachers and one who said she did not work with new teachers. All 83 of the coach respondents provided examples of how new teachers' needs differed from those of experienced teachers. Most responses fell into four categories: classroom management, core curriculum, modeling, and receptiveness.

Thirty-three percent of interviewed coaches (27 of 83) reported that new teachers had a greater need for help with classroom management, including skills such as establishing routines and procedures for students and managing student behavior.

Twenty-seven percent of interviewed coaches (22 of 83) said that new teachers needed more assistance with the core reading curriculum, which all K–3 teachers in Reading First were required to use at least 90 minutes a day. While 7 percent of coaches (6 of 83) said that experienced teachers also needed help with the core curriculum, all these coaches added that the type of support veteran teachers needed was different. For example, some veteran teachers needed more help refining their core program instruction, while new teachers needed help with fundamentals such as pacing.

Twenty-two percent of coaches (18 of 83) said that new teachers needed more modeling from the coach—a form of assistance in which the reading coach teaches a class while the classroom teacher observes. The coach might model strategies for increasing student engagement or speeding the pace of a lesson.

Coaches also commented on differences in teacher receptiveness to receiving help. During interviews, coaches sometimes spontaneously described new teachers as being easier to work with (28 percent; 23 of 83 coaches), more open (22 percent; 18 of 83), and more eager (19 percent; 16 of 83). In contrast,

TABLE 2

Frequency of coach feedback reported by new and experienced teachers, spring 2008

Teacher experience	Probability of reporting Once a month or more" (percent)	Odds	Odds ratio	95% confidence interval for odds ratio	*t* statistic	Degrees of freedom (approximate)	*p* value
New teachers	69.3	2.26	1.27	1.055, 1.525	2.53	5,260	.011
Experienced teachers	64.0	1.78					

Note: For this study's collapsed scale, 1 = once a month or more (original responses 3–6), and 0 = less than once a month (responses 1 and 2). The original scale was 1 = never, 2 = once or a few times a year, 3 = once a month, 4 = 2–3 times a month, 5 = 1–3 times a week, 6 = daily.

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

2 percent of coaches (2 of 83) said that new teachers were more difficult to work with than experienced teachers were. Comments about teacher receptiveness arose voluntarily as part of coaches' answers to a question about how teachers' needs differed. Since the question did not specifically ask about teacher receptiveness to coaching, not all coaches provided comments in this area.

Teachers' perceptions of the impact of coaching on instruction. When asked whether the coach had changed teachers' instruction during the past year, 85 percent of new teachers interviewed (72 of 85) and 61 percent of experienced teachers (51 of 83) said "yes." There were no differences between the descriptions by new and experienced teachers of how the coach had changed their instruction. For example, both new and experienced teachers described coaches as changing their instruction by providing resources, helping with data analysis and student grouping, and providing strategies for student engagement.

Five percent of interviewed new teachers (4 of 85), compared with 15 percent of interviewed experienced teachers (12 of 83), said the coach had not changed their instruction. The remaining new teachers (11 percent; 9 of 85) and experienced teachers (24 percent; 20 of 83) gave mixed or partial responses to this question. For example, these respondents said that the coach had only a minimal impact on their instruction this year or had been too busy to help or that they did not need coaching.

Both new and experienced teachers described coaches as changing their instruction by providing resources, helping with data analysis and student grouping, and providing strategies for student engagement

Teachers' perceptions of the support they received from their reading coach. To analyze teachers' overall perceptions of the support they received from their coach, a coach support scale was created from four survey items asking about whether the coach was an ally, a knowledgeable resource, or had helped teachers become more reflective (see table B1 in appendix B for list of all survey items). Responses to each of the four items were measured on a five-point scale (a higher score indicates a more positive perception of support). The four responses were averaged for each teacher to create a single variable.

The means for both groups were similar: 3.85 for new teachers and 3.81 for experienced teachers, which is close to "agree" (4.0) on the five-point scale (table 3). The differences between the two groups were not statistically significant.

How do teachers' experiences with collaboration differ?

This questions looks at regular teacher grade-level meetings, which are required in Reading First schools. Teacher survey and interview data were used to examine this question.

TABLE 3

New and experienced teachers' perceptions of support from their reading coach, spring 2008

Teacher experience	Mean on five point scale[a]	Difference between mean scores	95% confidence interval	*t* statistic	Degrees of freedom (approximate)	*p* value
New teachers	3.85	0.04	–0.03, 0.10	1.12	178	.263
Experienced teachers	3.81					

a. The estimated mean from the hierarchical linear model. Scale: 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly agree.

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

In interviews, 67 percent of new teachers and 74 percent of experienced teachers agreed that collaborative meetings were a good use of time

In a single survey item, teachers were asked, on a five-point scale, whether they agreed with the statement, "Attending grade-level team meetings is a good use of my time." A higher score meant more positive perceptions.

Analysis of the survey data showed that the mean response was 3.78 for new teachers and 3.79 for experienced teachers, or close to "agree" (4.0) on the five-point scale (table 4). Differences between new and experienced teachers were not statistically significant.

Interview data supported the survey findings. In interviews, 67 percent of new teachers (56 of 83 new teachers interviewed) and 74 percent of experienced teachers (59 of 80) agreed that the meetings were a good use of time. When asked to explain why, 64 percent of new teachers who agreed (36 of 56) and 29 percent of experienced teachers who agreed (17 of 59) said they appreciated the opportunity to collaborate. Additionally, 38 percent of new teachers who agreed (21 of 56) and 44 percent of experienced teachers who agreed (26 of 59) commented that the opportunity to review data and group students for instruction made the meetings worthwhile.

Interviewed teachers who reported not fully valuing grade-level team meetings said that they mostly valued meetings, but with some caveats: 25 percent of new teachers (21 of 83) and 19 percent of experienced teachers (15 of 80). Another smaller group reported not valuing meetings: 7 percent of new teachers (6 of 83) and 8 percent of experienced teachers (6 of 80). For both new and experienced teachers, a common explanation for not valuing the grade-level meetings was that they were unhelpful because of bickering, off-task conversations, or too much talking; of the teachers who did not fully value the meetings, 44 percent of new teachers (12 of 27) and 48 percent of experienced teachers (10 of 21) made similar comments.

How do teachers' perceptions of their data skills differ?

Responses to the question of whether new and experienced teachers' perceptions of their data skills differed were measured using five survey items, including confidence in using data to group students and understanding of school trends (see table B1 in appendix B for complete list of items). A higher score on the five-point scale meant greater confidence in skills. Responses to the five survey items were averaged for each teacher to create a single variable representing teachers' confidence in their data skills.

The mean confidence for both new and experienced teachers was at or above 4, or "agree," on the five-point scale. The mean confidence of new teachers (4.00) was lower than that of experienced teachers (4.20), and the difference was statistically significant (table 5).

TABLE 4

New and experienced teachers' views of the usefulness of collaborative meetings, spring 2008

Teacher experience	Mean on five point scale[a]	Difference between mean scores	95% confidence interval	*t* statistic	Degrees of freedom (approximate)	*p* value
New teachers	3.78	−0.01	−0.08, 0.06	-0.33	342	0.742
Experienced teachers	3.79					

a. The estimated mean from the hierarchical linear model. Scale: 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly agree.

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

How does teachers' support for the reform initiative differ?

Both teacher surveys and teacher interviews were used to assess teachers' support for the reforms brought to their school by Reading First.

Eight items on the survey related to whether teachers supported the instructional changes under Reading First—whether they liked the core program and trusted the assessments (see table B1 in appendix B for complete list of items). Responses to each item were measured on a five-point scale, with a higher score indicating a more positive perception of support. The eight responses were averaged for each teacher to create a single variable representing support for Reading First.

For both new and experienced teachers, mean scores for support for Reading First were close to the midpoint of the five-point scale, indicating that teachers "neither agreed nor disagreed": 3.08 for new teachers and 3.12 for experienced teachers (table 6). The difference was not significant.

Overall, 95 percent of new teachers and 95 percent of experienced teachers agreed or agreed with caveats that Reading First was good for them as a teacher

The interview data also showed no difference in support for Reading First. Overall, 95 percent of new teachers (81 of 85) and 95 percent of experienced teachers (78 of 82) agreed—or agreed with caveats—that Reading First was good for them as a teacher. For the 33 percent of new teachers (28 of 85) and 40 percent of experienced teachers (33 of 82) who offered caveats to their approval of the reform, concerns clustered around the scripted

TABLE 5
New and experienced teachers' confidence in their data skills, spring 2005

Teacher experience	Mean on five point scale[a]	Difference between mean scores	95% confidence interval	*t* statistic	Degrees of freedom (approximate)	*p* value
New teachers	4.00	−0.20	−0.25, −0.15	−8.55	2,200	<.001
Experienced teachers	4.20					

a. The estimated mean from the hierarchical linear model. Scale: 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly agree.

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

TABLE 6
New and experienced teachers' support for Reading First, spring 2008

Teacher experience	Mean on five point scale[a]	Difference between mean scores	95% confidence interval	*t* statistic	Degrees of freedom (approximate)	*p* value
New teachers	3.08	−0.04	−0.10, 0.03	−1.41	9,406	.157
Experienced teachers	3.12					

a. The estimated mean from the hierarchical linear model. Scale: 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly agree.

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

nature of the core curriculum and the difficulty that created for differentiating instruction to meet the needs of all students.

STUDY LIMITATIONS

The study findings should be considered in light of several limitations. First, there are limitations inherent in collecting survey and interview data, such as the tendency for some respondents to give consistently high or low ratings (over-rater or under-rater bias) and interviewer subjectivity (Isaac and Michael 1995). Second, because of the semistructured interview protocol, some respondents included information beyond the scope of the question, adding information on which not all respondents had a chance to comment. Third, there are limitations inherent in comparing new teachers with experienced teachers. The group of new teachers includes a portion that will leave teaching. Because these leavers may have characteristics that are different from those of teachers who stay in the field, new teachers may have different characteristics from experienced teachers. The data lacked information on all the demographic characteristics needed to address these differences, such as age, gender, and race/ethnicity. Fourth, the study is based on data from western states and so cannot be generalized to the entire United States without accounting for geographic and demographic differences. Similarly, experiences in Reading First may not be generalized to other reform initiatives.

Because this was a descriptive study, it is not possible to determine whether the differences found between new and experienced teachers had any impact on teacher retention or student learning. Nor is it possible to determine whether the differences found between new and experienced teachers are due to differences in years of teaching experience, since factors not accounted for in the study could have contributed to teachers' experiences and reactions to school reform (such as teachers' age, gender, and race/ethnicity). Finally, because the study analyzed existing data, some questions raised in the literature review (appendix A) could not be answered with the available data.

APPENDIX A
WHAT WE KNOW FROM THE LITERATURE

There is a large body of literature about new teachers, much of it focused on training, retention, and mentoring or induction programs. There is also a smaller body of literature about the elements of reform examined in this study: coaching, teacher collaboration, data use, and support for reforms. But there are very few studies of new teachers' experiences in reform initiatives or with specific elements of reform models (such as coaching, collaboration, and data use) and attitudes toward reform initiatives. This appendix describes the literature and its limitations.

Support from a coach

Coaching has become increasingly popular as a means of delivering job-embedded professional development (Russo 2004). Reading First schools hire literacy coaches to work with all K–3 teachers in the school. Coaches work with teachers in the classroom, observing, modeling, providing feedback, and planning lessons according to the needs of individual teachers. Coaching is expected to increase student achievement by improving teachers' instruction skills. However, there have been few large-scale studies of coaching and its impact on student achievement.

Studies have found considerable variation in how coaches do their jobs (Poglinco et al. 2003; Walpole and McKenna 2004). A recent study showed that Reading First coaches spent time on a variety of tasks including modeling instruction, assisting with testing, interpreting test data, observing teachers and giving feedback, and generally overseeing the Reading First program (Deussen et al. 2007). Kise (2006) suggests that such differentiated coaching methods may be appropriate to meet the varying needs of teachers, but the study does not provide any information about how the experience level of teachers might contribute to this differentiation.

One recent study of a statewide coaching initiative in Florida found that working with new teachers was a major emphasis for coaches and concluded that coaches working in schools with many new teachers could find themselves with an overwhelming caseload (Marsh et al. 2008). Unlike Reading First and other reform models, however, coaches in the Florida initiative were required to spend more time working with new teachers than with experienced teachers. There is little additional information about when and how coaches work with new teachers, and no information on the outcomes of that work.

Collaboration

Ending classroom teachers' isolation and increasing teacher collaboration is central to Reading First as well as other school reform initiatives. There is some evidence that teachers, regardless of years of experience, welcome collaboration. In a survey of teachers by the National Education Association (2003), teachers gave top ranking to "cooperative/competent teacher colleagues/mentors" for helping them teach well. Imbimbo (2004) reports that new teachers felt that support from colleagues was a top influence on their teaching efficacy. These studies looked at collaboration in its general sense, however, rather than collaboration in school reform settings.

For new teachers, collaboration is often one aspect of induction, which also can include mentoring, targeted professional development, and orientation. Because it is difficult to separate the impact of collaboration from that of other induction components, little is known about its impact on new teacher quality or retention (Lopez et al. 2004). However, Kardos (2004) finds a strong positive correlation between new teachers' regular collaboration with colleagues and their job satisfaction. Similarly, Smith and Ingersoll (2004) find that common planning time was associated with significantly lower turnover for new teachers These studies were correlational rather than experimental and therefore could not explore causal relationships.

The limited research on new teachers and collaboration has prompted a call for more research on

the circumstances under which new teachers support collaboration (Johnson, Berg, and Donaldson 2005).

Use of data

The use of student assessment data is now almost a universal expectation in school reform efforts. Scott (2007) finds that 97 percent of district officials in a nationally representative sample of Title I districts said that schools had increased the use of student achievement data in 2006/07 to inform instructional decisions in an effort to improve schools. In addition, the Comprehensive School Reform Quality Center (2005) reported that using data to make instructional decisions was a key element of 21 of 22 comprehensive school reform models used in almost 9,000 schools nationwide. The use of data in making instructional decisions is likely to continue, especially as part of the current movement toward tiered instruction initiatives such as response to intervention, which uses student assessment data to identify struggling students, choose academic interventions for these students, and monitor their progress (Griffiths et al. 2007).

While studies have shown a national trend in educators' use of data, particularly student achievement data, to make instructional decisions, use of the strategy varied. For example, Scott (2008) finds that all 42 principals in case study schools identified for restructuring under the No Child Left Behind Act of 2001 said use of data for instructional decisionmaking had increased in their school. However, their data use varied in frequency and intensity; some used data weekly to plan lessons while others used data only annually for student grouping.

There is little research, however, describing or measuring new teachers' use of data for instructional decisions. Recent studies suggest that many teachers, regardless of teaching experience, lack the data analysis skills necessary to use data to make sound instructional decisions (Dembosky et al. 2005; Marsh, Pane, and Hamilton 2006). One recent study of nationally representative teacher survey data found that even though data were increasingly available in schools, many teachers were not confident they had the training and tools to make data-driven decisions affecting instructional change (Gallagher, Means, and Padilla 2008). For example, 67 percent of teachers in schools that were not making adequate yearly progress on state tests believed that they could benefit from more professional development in adjusting instructional content based on student data.

Many studies of new teachers' use of data focus on their math and data skills, with an emphasis on their understanding of the subject area, rather than on their use of these skills for pedagogical purposes. For example, much of the research finds that preservice elementary education teachers tend to have low levels of math knowledge (Darling-Hammond 1999; Russell 1997; Scheaffer 2000) and difficulty interpreting statistical findings (Thomas 2002). These results suggest that new teachers generally do not enter the profession with a strong data orientation and may have trouble with the high level of data use increasingly required in school reform. This notion is supported by a study in four districts identified as leaders in data use; even these districts felt a need to provide training on data-based decisionmaking for new teachers (Datnow, Park, and Wohlstetter 2007).

Like other reform initiatives, Reading First expects teachers to use student data frequently and for multiple purposes. For example, teachers use data to group students, plan instruction, diagnose students' reading needs, monitor student progress, and understand student achievement trends across the school. This study compares new teachers' confidence in their data skills with that of their more experienced peers.

Support for the reform

Teacher support—or lack of support—for school reform can influence the likelihood of sustaining or scaling up the reform (Steiner 2000; Taylor 2005). Other reform initiatives acknowledge the importance of teacher support. For example, data

from the Comprehensive School Reform Quality Center (2005) show that efforts to increase teacher support for and fidelity to reforms was an element of 21 of 22 commonly used comprehensive school reform models.

While there is a small body of literature that explores teachers' attitudes toward school reform, little attention has been paid to how teachers' experience might affect reform. One study suggests that teachers' experience levels influenced their acceptance of and emotional responses to education reform initiatives (Hargreaves 2005). The study finds that teachers who were still early in their careers tended to be enthusiastic as well as adaptable and flexible, while experienced teachers were more skeptical of reforms and adapted only if they saw that a reform had long-term sustainability.

For Reading First, various state evaluations have found support for the reform among teachers (Autio, Roccograndi, et al. 2008; Autio, Scott, et al. 2008; Nelsestuen, Scott, and Burke 2008). The studies did not look at whether this support varied by experience, however, as the Hargreaves (2005) research suggests it might.

APPENDIX B
DATA SOURCES AND STUDY METHODS

Data for this study came from previous statewide evaluations of Reading First in six states (Alaska, Arizona, Idaho, Montana, Washington, and Wyoming), conducted in 2007/08 by the Northwest Regional Educational Laboratory (now Education Northwest) with funding from the state department of education in each state that did not require Office of Management and Budget (OMB) or Institutional Review Board (IRB) clearance.[6] State Reading First project staff members agreed to share evaluation data for this study, which also included previously collected teacher surveys and teacher and coach interview data (see below). Since the study involved only secondary analysis of existing data (no original data were collected), OMB or IRB clearance was not required.

Data sources

This study drew on existing teacher survey, teacher interview, and coach interview data.

Teacher surveys. A team of evaluators at Northwest Regional Educational Laboratory developed the Reading First surveys as part of statewide evaluations. These paper-and-pencil surveys were mailed to Reading First coaches at all 235 Reading First schools to administer to teachers over a three-week period in spring 2008. Teachers were instructed to complete their survey independently and received envelopes in which to seal their responses to ensure confidentiality. Surveys were mailed back in a single packet by the reading coach at each school. Evaluators followed up with schools that did not respond by the due date, and Reading First state directors generally encouraged schools to participate, but there were no programmatic or other repercussions for not returning surveys.

A subset of 19 survey items in four categories was used in this study, common across all six states: support from the reading coach, teacher collaboration, data confidence, and support for the reform (see table B1). Factor analysis of the 19 items confirmed that the four areas of the study were conceptually distinct.

Factor analysis was used to group like survey items and create scales representing these items. Principal component factor analysis was used to determine whether scales could be created from the survey items related to three of the topics of interest—coach support, data confidence, and support for reform. The factor analysis used a promax rotation because this technique does not require that the underlying variables be uncorrelated (Hair et al. 1998)—teachers' experiences with and perceptions of various components of Reading First are probably closely related to one another. With factor analysis, it is important to check the resulting dimensions (scales) for internal consistency; this was done using Cronbach's alpha. Dimensions with alpha of 0.6 or greater reliability were considered, as recommended by Hair et al. (1998). Three dimensions emerged that accounted for 62 percent of the variance. Table B2 shows how the survey items loaded onto each dimension.

Once these dimensions were identified, a scale score was created for each teacher by averaging the teacher's responses to the items in each dimension. The three variables created were then used in subsequent analyses.

Two additional variables were used in this study. First, the item "This year, how often did the reading coach provide you with feedback on your instruction?" was used individually because it had a different response scale than other items (see table B1). Second, the item, "Attending grade-level meetings is a good use of my time" was used for usefulness of collaboration because there were no other questions about collaboration. Both items were considered important to the study because they measured concepts central to the research questions. Prior to analysis, the distribution of responses to these two items was examined to determine whether they were appropriate to include in the hierarchical linear modeling.

Survey response rates were high: 98 percent of schools returned surveys. The overall teacher response rate was 91 percent, with a range of 85–98 percent in each state (table B3).

TABLE B1
Teacher survey items, spring 2008

Research area	Survey items[a]
Coach support (a variable for coach support was created by averaging four survey items)	Our reading coach has helped me become more reflective about my teaching practice.
	Even when providing critical feedback, I feel our reading coach is an ally in helping me to improve my instruction.
	Our reading coach is a knowledgeable resource about reading research and practices.
	Our reading coach has increased my understanding of how children learn to read.
Frequency of coach feedback	This year, how often did the reading coach provide you with feedback on your instruction?[b]
Data confidence (a variable was created by averaging five survey items)	I am confident in my ability to use data to plan small-group instruction.
	I am confident in my ability to use data to group students.
	I am confident in my ability to diagnose a student's specific reading needs using reading-assessment data.
	I am confident in my ability to understand student achievement trends across our school.
	I am confident in my ability to administer progress monitoring assessments.
Usefulness of collaborative meetings	Attending grade-level meetings is a good use of my time.
Support for reform (a variable was created by averaging eight survey items)	I strongly support the instructional changes that are occurring under Reading First.
	I am pleased that our school has a Reading First grant.
	I have significant philosophical or pedagogical objections to the approach of Reading First.[c]
	I think the DIBELS is a valid, accurate indicator of student reading ability.
	In my view, Reading First overemphasizes the importance of using DIBELS results.[c]
	I am very satisfied with the core reading program we are using at our school.
	I feel that I have a voice in our school's decision making about Reading First.
	Instruction in other subjects has suffered because of all of the focus on Reading First.[c]

a. Unless otherwise stated, all survey items had the following scale: 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly disagree.

b. This item was measured using the following scale: 1 = never, 2 = once or a few times a year, 3 = once a month, 4 = two to three times a month, 5 = one to three times a week, and 6 = daily.

c. The scale on these items was reversed for analysis, since the statements are negative, whereas other statements are positive.

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

The original dataset of the survey of Reading First teachers consisted of the responses of 3,094 K–3 teachers, 2,991 of whom reported their years of experience (table B4).

Coach and teacher interviews. In spring 2008, coach and teacher interview data were collected from 85 (36 percent) of the Reading First schools across the six states as part of the statewide evaluations conducted by Northwest Regional Education Laboratory (table B5 lists the interview items for the semistructured interviews with coaches and teachers). The process for selecting the schools for qualitative data collection varied across the six states, depending on the evaluation needs and preferences of the state project staff as well as the resources allocated for site visits in each state (table B6). In Wyoming, all seven schools were visited. In the other five states, schools were randomly selected for visits. These random selections were sometimes stratified by district (Alaska), by geographic region (Arizona and Idaho), or by cohort (Washington). In Montana, a random selection from all the schools was possible.

TABLE B2
Scales from survey items by research area, spring 2008

Scale	Item	Factor loading	Cronbach s alpha	Mean	Standard deviation
Coach support	Our reading coach has helped me become more reflective about my teaching practice.	.929	.921	3.8	.92
	Even when providing critical feedback, I feel our reading coach is an ally in helping me to improve my instruction.	.923			
	Our reading coach is a knowledgeable resource about reading research and practices.	.910			
	Our reading coach has increased my understanding of how children learn to read.	.880			
Data confidence	I am confident in my ability to use data to plan small-group instruction.	.868	.848	4.1	.62
	I am confident in my ability to use data to group students.	.854			
	I am confident in my ability to diagnose a student's specific reading needs using reading-assessment data.	.830			
	I am confident in my ability to understand student-achievement trends across our school.	.736			
	I am confident in my ability to administer progress-monitoring assessments.	.676			
Support for reform	I strongly support the instructional changes that are occurring under Reading First.	.797	.846	3.1	.77
	I am pleased that our school has a Reading First grant.	.768			
	I have significant philosophical or pedagogical objections to the approach of Reading First.[a]	.747			
	I think the DIBELS is a valid, accurate indicator of student reading ability.	.706			
	In my view, Reading First overemphasizes the importance of using DIBELS results.[a]	.702			
	I am very satisfied with the core reading program we are using at our school.	.637			
	I feel that I have a voice in our school's decision making about Reading First.	.598			
	Instruction in other subjects has suffered because of all of the focus on Reading First.[a]	.592			

Note: Survey items had the following scale: 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree

a. The scale on these items was reversed for analysis, since the statements are negative, whereas other statements are positive.

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

The different sampling plans in the six states meant that the states were differentially represented in the qualitative database. This issue was addressed during a preliminary analysis by categorizing responses by state to determine whether findings were state specific. Results did not differ substantially by state, and so state was not considered in the final analysis of the qualitative data.

TABLE B3

Teacher survey response rates, by state, spring 2008

State	Number of schools surveyed	School level response rate for teacher surveys (percent)	Number of K 3 teachers surveyed	Teacher level response rate for teacher surveys (percent)	
				Number	Percent of total
Alaska	14	100	113	107	95
Arizona	85	100	1,339	1,291	96
Idaho	32	97	486	442	91
Montana	12	95	125	123	98
Washington	85	96	1,254	1,064	85
Wyoming	7	100	72	67	93
Total	235	98	3,389	3,094	91

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

TABLE B4

Number of teacher survey respondents at two levels of teaching experience, by state, spring 2008

State	New teachers (1 4 years)	Experienced teachers (5 or more years)
Alaska	22	81
Arizona	519	729
Idaho	141	297
Montana	27	96
Washington	288	726
Wyoming	19	46
Total	1,016	1,975

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

The reading coach was interviewed at each of the 85 schools. All but one site visit also included interviews with two teachers, for a total of 169 teacher interviews (one Washington site visit included only one teacher interview).

In all six states, the newest teacher at each school was interviewed. In five states, the second interviewed teacher was selected from all remaining K–3 teachers by choosing the teacher whose first name came last in the alphabet. This was slightly different in Arizona, a state that emphasized Reading Leadership Teams in Reading First. The second teacher in Arizona was the Reading Leadership Team member with the most years of teaching experience. Using this process, 71 of the 85 schools included interviews with both a new teacher (one to four years of experience) and an experienced teacher (five or more years of experience). At six schools, two experienced teachers were interviewed since the newest K–3 teacher had at least five years of experience. At another seven schools, two new teachers were interviewed because the second teacher selected alphabetically was a new teacher. At the final school, only one teacher was able to be interviewed (a new teacher). In total, there were 85 interviews with new teachers and 84 interviews with experienced teachers.

All respondents were promised confidentiality; neither their name nor the name of their school was linked to their responses in any of the evaluation reports or in this study. In addition, care was taken not to provide descriptive information that might identify participants. For example, no quotations or individual responses were used if they included contextual clues that could be used to identify a respondent. Notes from the interviews were stored in a database that was password protected. The interview notes were accessible only to the person who conducted the interview, the project coordinator and, later, the evaluators and researchers who analyzed the data.

TABLE B5
Interview items for coaches and teachers, by research area, spring 2008

Research area	Respondent	Interview items
Coach support	Coach	Tell me about working with inexperienced teachers this year, particularly those with 1–4 years or fewer of experience. Was this part of your role? Do new teachers have different needs than veteran teachers? Please describe.
Coach support	Teacher	Has your coach helped you change your instruction this year? If so, how? (Please provide an example.)
Usefulness of collaborative meetings	Teacher	Do you think that attending grade-level team meetings is a good use of your time? Why or why not?
Support for reform	Teacher	To what degree is Reading First good for you as a teacher? Why?

Source: Authors' analysis of the 2008 teacher and coach interviews administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) the statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

TABLE B6
Interview sample size and selection method, by state, spring 2008

	Reading First schools selected for interviews		
State	Number	Percent of total	Selection method
Alaska	6	43	Random selection of two schools within each of three districts
Arizona	28	34	Random selection of schools, stratified within seven geographic regions
Idaho	15	47	Random selection of schools stratified within three geographic regions
Montana	6	50	Random selection of half the fully funded schools
Washington	23	27	Random selection of half the cohort 3 (newest) schools and 12 of 69 more experienced schools
Wyoming	7	100	All fully funded cohort 2 and 3 schools selected
Total	85	36	na

na is not applicable.

Source: Authors' analysis of the 2008 teacher and coach interviews administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

Interviewers were experienced evaluators and received two-day training in the interview protocols. Both the teacher and coach protocols were semistructured. Interviewers were instructed to ask each question as it appeared on the protocol and then to follow up with probes to clarify any responses that did not include enough information or answer the question. These protocols were based on the stated aims of Reading First in national and state documents, interviews with state Reading First project staff, and findings of past evaluation reports. The protocol for teachers included three questions relevant to this study; one coach interview question was analyzed for this study (see tables B5 and B7).

Evaluators also collected information on teachers' years of teaching experience and the state they worked in. The number of responses to each question varied, as teachers occasionally did not address the question or chose not to respond (see table B7).

Interviews with coaches and teachers were not recorded. Evaluators were asked to take detailed notes, as close to verbatim as possible.

Sample

The six states in the study included 235 fully funded Reading First grantees in the 2007/08

TABLE B7

Total number of interview respondents for each question, by level of experience, spring 2008

Respondent	Item	New teachers	Experienced teachers	Total
Teacher	Has your coach helped you change your instruction this year? If so, how? (Please provide an example.)	85	83	168
	Do you think that attending grade-level team meetings is a good use of your time? Why or why not?	83	80	163
	To what degree is Reading First good for you as a teacher? Why?	85	82	167
Coach	Do new teachers have different needs than veteran teachers? Please describe.	na	na	85

na is not applicable.

Source: Authors' analysis of the 2008 teacher and coach interviews administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

school year.[7] All 235 schools were included in the survey dataset. By locale, 33 percent of schools were urban, 30 percent were rural, 26 percent were in an urban fringe, and 10 percent were in small or large towns (table B8). Some 46 percent of schools served more than 500 students, and 22 percent served 300 or fewer students.[8] Schools had been in Reading First for one year (4 percent), two years (44 percent), three years (5 percent), four years (17 percent), or five years (30 percent). Coach and teacher interviews were collected from a subset of 85 randomly selected schools. The available characteristics of the teachers from the participating schools in the study are shown in table B9.[9]

Data analyses

This is a descriptive study with a concurrent mixed methods design (Creswell 2003). Analyses of teacher surveys, teacher interviews, and coach interviews, described below, were completed concurrently.

Survey analyses. Survey data were analyzed using hierarchical linear models and, for one survey item, a hierarchical generalized linear model, to address the nesting structure that existed in the data (teachers nested within schools). Analysis of missing data was performed prior to the analysis with hierarchical linear models, to address the incompleteness of the dataset.

TABLE B8

Characteristics of participating schools, 2003/04

Characteristic	Percentage of schools (n = 233[a])
School location	
Urban	33
Urban fringe	26
Small/large town	10
Rural	30
School size	
<100 students	6
101–300 students	16
301–500 students	32
>500 students	46
Years in Reading First	
1	4
2	44
3	5
4	17
5	30
Percentage of students from 233 schools (n = 110,860)	
Poverty	
Eligible for free or reduced-price lunch	74
Ethnic/racial composition	
White	31
Hispanic	53
African American	6
Native American	7
Asian	4

a. Data are not available for 2 of the 235 schools.

Source: Authors' analysis based on U.S. Department of Education (2006) and, for years in Reading First, on data collected during Northwest Regional Educational Laboratory's statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

Missing-data analysis. In preparing to analyze how a teacher's experience predicted teacher responses to the scales and items described above, missing data were examined to see whether the nature of the missing data might bias the results. After cleaning, the data file contained survey results from 3,094 teachers. Of these, 213 had missing data in one or more of the variables of interest (7 percent). Since this value exceeded the pre-set cutoff for listwise deletion (5 percent), multiple imputations were performed to produce three complete and independent datasets (each with $N = 3{,}094$) using NORM software (Schafer 1999). Each of the three datasets was then used in the hierarchical linear model analysis to produce three independent estimates of population parameters and their standard errors. Finally, the separate estimates were pooled to produce a single set of population parameters, their standard errors, and their confidence intervals using Rubin's (1987) rules for scaler estimands.

Hierarchical linear model analysis 1: estimating the intraclass correlation coefficient. Because of the nested nature of the data—teachers are nested within schools—the next stage of the analysis was thought to entail hierarchical linear modeling for four of the five survey items/scales measured with a five-point Likert scale (Hox 2002; Kreft and de Leeuw 1998; Raudenbush and Bryk 2002). Analysis of the intraclass correlation coefficient (ICC) confirmed that belief. The ICC represents the proportion of the total variance in the outcome between schools rather than within a school. Mathematically, it is expressed as $ICC = \tau / (\tau + \sigma^2)$, where τ is the variance in the teacher response at the school level and σ^2 is the variance at the teacher level. ICC values range from 0 to 1. A large ICC (such as 0.30) means that a substantial portion of the variation among teacher responses is associated with the school in which they teach (teachers within the same school answer in fairly similar ways). A small ICC (such as 0.01) indicates that little of the variation among individual teachers is associated with the school. When the ICC is extremely small, ignoring the nesting and using a single-level model will yield reasonably accurate results.

The following hierarchical linear model, a one-way analysis of variation (ANOVA) model, was fit to the data to calculate τ, the variance of u_{0j}, and σ^2, the variance of r_{ij} for each of the variables (results are in table B10):

Level 1 (teacher level) model

$$Y_{ij} = \beta_{0j} + r_{ij}$$

Level 2 (school level) model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

The ICC was then hand-calculated for each variable using the equation in the previous paragraph.

Across the measures, the ICC was large enough to merit the use of a multilevel modeling technique to account for the clustering of teacher responses at the school level.

Hierarchical linear model analysis 2: estimating the strength of association between teacher experience and teacher response. After the ICCs were calculated, the following hierarchical linear model was used to estimate the strength of association between the degree of teacher experience and teacher response:

Level 1 (teacher level) model

$$Y_{ij} = \beta_{0j} + \beta_{1j}[\text{New}] + r_{ij}$$

Level 2 (school level) model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$

where *Yij* is teacher response (individual or scale items from table B10), and γ_{10} is the estimated true difference between new teachers (one to four years of experience) and experienced teachers (five or more years of experience). Because coding was done using the experienced teachers as the referent category, the coefficient associated with being either a new or experienced teacher (β_{1j} for

TABLE B9

Characteristics of surveyed and interviewed teachers, spring 2008

	Surveyed teachers		Interviewed teachers	
Characteristic	New	Experienced	New	Experienced
Total number	1,016	1,975	85	84
Average years of experience (standard deviation)	2.3 (1.1)	15.3 (8.8)	1.9 (1.1)	15.2 (7.8)
Master's degree or above (percent)	223 (21)	1,018 (51)	na	na
Grade taught in 2007/08 (percent)				
Kindergarten	24	21	24	14
1	26	23	29	23
2	24	24	32	33
3	21	22	15	30

na is not applicable.

Source: Authors' analysis of the 2008 teacher surveys and interviews collected as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

TABLE B10

Results of random analysis of variance model for calculating the intraclass correlation coefficient

Scale/item	Unconditional intraclass correlation coefficient
Coach support	0.23
Usefulness of collaborative meetings	0.12
Data confidence	0.06
Support for reform	0.23

Source: Authors' analysis of the 2008 teacher surveys administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

each school and γ_{10} for the population of schools) had a negative value when the estimated mean for experienced teachers was higher than that for new teachers. A fixed-slope model ($\beta_{1j} = \gamma_{10}$) was used because the number of new teachers per school was often too small to produce a reliable estimate of random slopes.

Hierarchical generalized linear model analysis 1. To calculate the ICC for frequency of coach feedback, the following hierarchical generalized linear model was used:

Level 1 (teacher level) model

$$\text{Sampling model: } y_{ij}|\varphi_{ij} \quad \sim \text{Bernoulli}(\varphi_{ij})$$

$$\text{Link function and structural model: } \log[\varphi_{ij} / (1 - \varphi_{ij})] = \beta_{0j}$$

Level 2 (school level) model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

where y_{ij} represents a teacher response of either "Once a month or more" or "Less than once a month." This dichotomous variable is assumed to have a Bernoulli distribution.

The level 1 model consists of the Bernoulli sampling model, a logit link function to "linearize" the probability (φ_{ij}), and a structural model for the linearized outcome. The structural models for level 1 and 2 do not include any covariates, so that level 2 residual variance is unconditional and is used to calculate the unconditional ICC. Several equations could be used to calculate the ICC under binomial hierarchical generalized linear modeling; the simplest is based on a threshold model, with ICC expressed as $ICC = \tau / (\tau + \pi 2/3)$, where τ is the variance of the level 2 residual and π is a constant 3.141593.

The independent estimates of τ from the three imputed datasets obtained with this unconditional model were used to calculate an ICC for frequency of coach feedback of 0.301, meaning that 30 percent of the variance in the teacher response to this item was associated with the difference in schools. (Technically, 30 percent of the variance in the log-odds of the expected probability of a teacher answering "Once a month or more" resides at the school level.)

As the variance of the level 2 residual, τ represents the degree of heterogeneity across schools in terms of expected teacher response to the item asking about frequency of coach feedback. The variance component analysis showed a significant τ; hence, the necessity for including the school random effect in the model.[10]

Hierarchical generalized linear model analysis 2: estimating the strength of association between teacher experience and teacher response. The analysis of frequency of coach feedback used hierarchical generalized linear modeling with a logit link function to estimate the strength of the association between teacher experience and teacher response. In the model described below, y_{ij} represents a teacher response ("Once a month or more" or "Less that once a month"), a dichotomous variable assumed to have Bernoulli distribution:

Level 1 (teacher level) model

$$\text{Sampling model: } y_{ij}|\varphi_{ij} \quad \sim \text{Bernoulli}(\varphi_{ij})$$

Link function and structural model:

$$\log[\varphi_{ij} / (1 - \varphi_{ij})] = \beta_{0j} + \beta_{1j}[\text{New}]$$

Level 2 (school level) model

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

The level 1 model consists of the Bernoulli sampling model, a logit link function to linearize the probability, and a structural model for the linearized outcome. The term γ_{10} represents the estimated true difference between new teachers (one to four years of experience) and experienced teachers (five or more years of experience) on the log-odds metric. Because coding was done using the experienced teachers as the referent category, the coefficient associated with being either a new or experienced teacher (β_{1j} for each school and γ_{10} for the population of schools) had a positive value when being a new teacher was associated with an increased odds or probability of answering "Once a month or more." A fixed slope model ($\beta_{1j} = \gamma_{10}$) was used because the number of new teachers per school was often too small to produce a reliable estimate of random slopes.

Once the gamma coefficients were estimated, the odds were calculated by adding the coefficients while still on the log-odds (linearized) metric and then taking an exponential of this value. The following formula was used to transform odds (O) to a probability (φ_{ij}) of reporting "Once a month or more": $\varphi_{ij} = O / (1 + O)$. All these calculations were done by hand. A unit-specific model was used for the analysis since interest was in the association between new teacher status and a teacher response in a given school.

Interview analyses. Interview data from teacher and coach interviews were loaded into Atlas-ti Version 5.0 from Scientific Software Development to facilitate coding and analysis. All data were located in a secure file accessible to researchers. Initially, two of the authors developed a code structure that included a list of "start codes" keyed to the interview questions. For example, teachers' answers to the question, "Has your coach helped you change your instruction this year? If so, how?" was initially coded simply "yes," "no," or "partial" without further qualifying the themes within the responses. Each researcher checked the utility of these codes on half the total set of data. The start codes are listed in table B11.

Coding took place in three stages—open coding, coding a common set of responses, and final coding. In the open coding, researchers developed

TABLE B11
Interview data analysis code structure

Code type	Code and responses				
	Follow core program?	Expectations reasonable?	Coach changed instruction?	Grade level meetings good use of time?	Reading First good for you as a teacher?
Start codes	FP100% FPMostly FPExceptions	FPYes FPNo FPSome	CHYes CHNo CHPartial	MTGYes MTGNo MTGPartial	RFGHigh RFGMiddle RFGLow
Inductive codes, teachers	Support from coach	Use of data	Collaboration	Other	
	Coaching	Data	Collaboration	Core	Fidelity
	Coach teaches students	Accountability	Lesson planning	Alternate core	Change
	Coach not in class	Results	Map-template	Supplements	Constraints
	Modeling	Differentiation	Time	Interventions	Overwhelming
	Advice	Walk to read	Be consistent	High students	Other subjects shortchanged
	Pacing	State tests	Strategies	Low students	Structure
	Engagement	Fluency		English language learner students	Professional development
	Observations	Placements		Supports students	Funding
	Organize for instruction	Assessment		Time reading	Better teacher
	Feedback	Progress monitor			Paraprofessionals
	Classroom management				Limited prep
	Grouping				Learned about reading
	Resource				Reading skills missing
	Affective				
Inductive codes, coaches	Working with new teachers	Coaching focus new teachers	Approaches new teachers	Working with experienced teachers	Focus/approach experienced teachers
	Eager	Affective	Meetings	Experienced attitude	Experienced adjust to students' needs
	Easier	Basics	Modeling	Experienced change	Experienced classroom management
	Harder	Classroom management	Observations	Experienced collaboration skills	Experienced content background
	Same	Curriculum	Fidelity	Experienced convinced by data	Experienced curriculum
	Take less time	Engagement	Training sessions	Experienced do their own thing	Experienced instruction
	Take more time	Organize for instruction		Experienced go deeper	Experienced engagement
	More open	Instruction		Experienced harder	Experienced observations

(CONTINUED)

TABLE B11 (CONTINUED)
Interview data analysis code structure

Code type	Code and responses				
	Follow core program?	Expectations reasonable?	Coach changed instruction?	Grade level meetings good use of time?	Reading First good for you as a teacher?
Start codes	FP100% FPMostly FPExceptions	FPYes FPNo FPSome	CHYes CHNo CHPartial	MTGYes MTGNo MTGPartial	RFGHigh RFGMiddle RFGLow
Inductive codes, coaches (continued)	Challenge Reading First	Lesson planning		Experienced know students	Experienced pacing
	Coaching challenge	Maps-templates		Experienced low expectations	
	Overwhelmed	Orienting		Experienced more experience	
	Know it all	Pacing		Experienced more skills	
	Sensitive	Progress monitor		Experienced new perspective	
	Supported by other	Using data		Experienced take more time	
	College preparation				
	Content knowledge				

Note: Although codes are grouped by relationship with the research questions, codes were often used across interview questions. For example, teachers may have mentioned "data" in response to several questions.

Source: Authors' analysis of the 2008 teacher and coach interviews administered as part of Northwest Regional Educational Laboratory's (now Education Northwest) statewide evaluations of Reading First in Alaska, Arizona, Idaho, Montana, Washington, and Wyoming.

codes inductively based on the content of the qualitative responses. In this stage of the analysis, each researcher coded half the data using the start codes and then developed a list of inductive codes. Qualitative researchers then discussed the operational definitions of these inductive codes with the full research team to resolve any potentially problematic codes and finalize the list of codes. The final inductive codes are listed in table B11.

After the codes were finalized, the two researchers coded a common set of responses (a representative sample of approximately 25 percent of the interview data) using the inductive codes to determine interrater reliability and reach consensus on how codes would be operationally defined. This process continued with an additional 5 percent of the data until researchers were coding with 98 percent agreement.

To complete the final coding, the two researchers split all the data again by states. Each researcher took the portion of the data that the other had coded during the open coding and completed the coding using the final inductive codes. Thus, each researcher read all the interview data. As a final check, researchers reviewed one another's coding, and the few minor discrepancies were easily handled.

Next, data displays were created that addressed the four research questions (Miles and Huberman 1994). These tables were sorted by interview protocol questions to determine how responses to the interview questions related to the research study questions. Tables were also sorted by teacher experience level to determine how responses varied by teacher experience. Finally, tables were sorted by state to ensure that any responses that varied

by state could be reported, although variations by state were not expected.

Written memos illustrated typical comments and the range of remarks related to findings. In addition, data displays, memos, and queries (Miles and Huberman 1994) conducted within Atlas-ti created further opportunities to look for confirming and disconfirming evidence. These steps allowed the research team to develop its findings in preparation for writing the final report.

NOTES

1. The Comprehensive School Reform Quality Center (2005) report contained descriptions of each reform model. The counts cited here were obtained by systematically searching the descriptors in the report and summing how many models included coaching, collaboration, data use, and buy-in as part of their model.

2. In the response to intervention tiered instruction approach, students first receive support in general education classrooms through the core program. For students who do not respond to classroom instruction, interventions of increasing intensity are added. The goal is to improve education outcomes by identifying struggling students, providing them with targeted interventions, and monitoring responses to interventions.

3. This definition of "new" and "experienced" teachers was based on research findings that indicate that the steep learning curve for new teachers extends well beyond the first year of teaching (Steffy and Wolfe 1997; Steffy et al. 2000). Teachers spend several of their early career years increasing their knowledge of pedagogy and becoming confident teachers.

4. The six responses options were 1 = never, 2 = once or a few times a year, 3 = once a month, 4 = two to three times a month, 5 = one to three times a week, and 6 = daily. Statewide evaluation reports from the six states in this study have collapsed this item into the same dichotomous variable for reporting; see Nelsestuen, Scott, and Burke (2008) and Autio, Roccograndi, et al. (2008) as examples.

5. Converting probability [φ] to odds [$\varphi / (1 - \varphi)$] yielded odds for new teachers of 2.26, or 1.27 times that for experienced teachers (1.78). The increase was statistically significant.

6. In Arizona, Northwest Regional Educational Laboratory partnered with the Arizona Prevention Resource Center for the evaluation.

7. Some states had additional Reading First schools that were not fully funded. These schools finished their full grants after three years and received partial funding and/or state assistance in 2007/08. These schools were not included in this study.

8. Average school size differed by state. Alaska, Montana, and Wyoming had smaller average school sizes than Arizona, Idaho, and Washington.

9. Teacher characteristics such as gender and race were not available as they were not collected by evaluators. The only characteristic about interviewed teachers that was collected was their years of experience.

10. The result of variance component analysis for each of the imputed data sets was as follows: for dataset 1, $\tau = 1.397$, $\chi^2(224) = 943.3$, $p < 0.001$; for dataset 2, $\tau = 1.404$, $\chi^2(224) = 943.6$, $p < 0.001$; for dataset 3, $\tau = 1.385$, $\chi^2(224) = 938.9$, $p < 0.001$.

REFERENCES

Arizona Resource Prevention Center. (2008). *Arizona Reading First annual external evaluation report, cycle 2, year 6 2007–2008.* Tempe, AZ: Arizona State University.

Autio, E., Roccograndi, A., Nelsestuen, K., Burke, A., Greenberg-Motamedi, J., Deussen, T., and Davis, A. (2008). *Washington Reading First evaluation report 2007–2008.* Portland, OR: Northwest Regional Educational Laboratory.

Autio, E., Scott, C., Burke, A., and Stewart, V. (2008). *Wyoming Reading First annual evaluation report 2007–2008*. Portland, OR: Northwest Regional Educational Laboratory.

Clotfelter, C., Ladd, H., Vigdor, J., and Wheeler, J. (2007). *High-poverty schools and the distribution of teachers and principals* (CALDER Working Paper 1). Washington, DC: Urban Institute, National Center for Analysis of Longitudinal Data in Education Research. Retrieved October 1, 2008, from www.caldercenter.org/PDF/1001057_High_Poverty.pdf.

Comprehensive School Reform Quality Center (2005). *CSRQ Center report on elementary school comprehensive school reform models.* Washington, DC: American Institutes for Research.

Creswell, J.W. (2003). *Research design: qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Darling-Hammond, L. (1999). Educating teachers. *Academe, 85*(1), 26–33.

Datnow, A., Park, V., and Wohlstetter, P. (2007). *Achieving with data: how high-performing school systems use data to improve instruction for elementary students.* Los Angeles, CA: University of Southern California, Center on Educational Governance.

Davis, A., and Greenberg-Motamedi, J. (2008). *Alaska Reading First annual evaluation report 2007–2008.* Portland, OR: Northwest Regional Educational Laboratory.

Dembosky, J.W., Pane, J.F., Barney, H., and Christina, R. (2005). *Data driven decisionmaking in southwestern Pennsylvania school districts.* Working paper. Santa Monica, CA: RAND.

Deussen, T., Coskie, T., Robinson, L., and Autio, E. (2007). *"Coach" can mean many things: five categories of literacy coaches in Reading First* (Issues & Answers Report, REL 2007–No.005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Deussen, T., Nelsestuen, K., and Scott, C. (2008). *What have we learned after five years? Lessons from NWREL's state evaluations of Reading First.* Portland, OR: Northwest Regional Educational Laboratory.

Gallagher, L., Means, B., and Padilla, C. (2008). *Teachers' use of student data systems to improve instruction: 2005 to 2007.* Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service. Retrieved October 26, 2008, from www.ed.gov/rschstat/eval/tech/teachers-data-use-2005-2007/teachers-data-use-2005-2007-intro.html.

Gamse, B.C., Bloom, H.S., Kemple, J.J., Jacob, R.T., Boulay, B., Bozzi, L., Caswell, L., Horst, M., Smith, W.C., St. Pierre, R.G., Unlu, F., Herlihy, C., and Zhu, P. (with Greene, D., Kim, J., LaLiberty, D., Lam, K., Maree, K., McCormick, R., Miura, J., Unterman, R., and Wong, E.). (2008). *Reading First impact study: interim report* (NCEE 2008–4016). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Griffiths, A.J., Parson, L.B., Burns, M.K., VanDerHeyden, A., and Tilly, W.D. (2007). *Response to intervention: research for practice.* Alexandria, VA: National Association of State Directors of Special Education.

Hair, J.F., Jr., Anderson, R.E., Tatham, R.L., and Black, W.C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Hargreaves, A. (2005). Educational change takes ages: life, career, and generational factors in teachers' emotional responses to educational change. *Teaching and Teacher Education, 21*(8), 967–83.

Hox, J. (2002). *Multilevel analysis: techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Imbimbo, J. (2004). *The voice of the new teacher.* Washington, DC: Public Education Network. Retrieved March 20, 2009, from www.publiceducation.org/pdf/Publications/Teacher_Quality/Voice_of_the_New_Teacher.pdf.

Isaac, S., and Michael, W.B. (1995). *Handbook in research and evaluation*: *a collection of principles, methods, and strategies useful in the planning, design, and evaluation of studies in education and the behavioral sciences* (3rd ed.). San Diego, CA: Edits.

Johnson, S.M., Berg, J.H., and Donaldson, M.L. (2005). *Who stays in teaching and why: a review of the literature on teacher retention.* Cambridge, MA: Harvard Graduate School of Education, Project on the Next Generation of Teachers. Retrieved October 22, 2007, from http://assets.aarp.org/www.aarp.org_/articles/NRTA/Harvard_report.pdf.

Kardos, S. (2004). *Supporting and sustaining new teachers in schools: the importance of professional culture and mentoring.* Cambridge, MA: Harvard University Press.

Kise, J.A.G. (2006). *Differentiated coaching: a framework for helping teachers change.* Thousand Oaks, CA: Corwin Press.

Kreft, I., and de Leeuw, J. (1998). *Introducing multilevel modeling.* Thousand Oaks, CA: Sage.

Lankford, H., Loeb, S., and Wycoff, J. (2002). Teacher sorting and the plight of urban schools: a descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37–62.

Lopez, A., Lash, A., Schaffner, M., Shields, P., and Wagner, M. (2004). *Review of research on the impact of beginning teacher induction on teacher quality and retention.* Menlo Park, CA: SRI International. Retrieved October 22, 2007, from: http://policyweb.sri.com/cep/publications/Induction_LitReviewFeb_2004.pdf.

Marsh, J.A., McCombs, J.S., Lockwood, J.R., Francisco, M., Gershwin, D., Naftel, S., Le, V., Shea, M., Barney, H., and Crego, A. (2008). *Supporting literacy across the sunshine state: a study of Florida middle school reading coaches.* Santa Monica, CA: RAND.

Marsh, J.A., Pane, J.F., and Hamilton, L.S. (2006). *Making sense of data-driven decision making in education: evidence from recent RAND research.* Santa Monica, CA: RAND.

Miles, M.B., and Huberman, A.M. (1994). *Qualitative data analysis: an expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

National Education Association. (2003). *Status of the American public school teacher 2000–2001.* Retrieved March 20, 2009, from www.nea.org/home/2233.htm.

Nelsestuen, K., Scott, C., and Burke, A. (2008). *Idaho Reading First 2007–2008 evaluation report*. Portland, OR: Northwest Regional Educational Laboratory.

Northwest Regional Educational Laboratory. (2008). *Reading First site visitor training materials*. Portland, OR: Northwest Regional Educational Laboratory.

Poglinco, S.M., Bach, A.J., Hovde, K., Rosenblum, S., Saunders, M., and Supovitz, J.A. (2003). *The heart of the matter: the coaching model in America's Choice schools.* Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Raudenbush, S.W., and Bryk, A.S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Roccograndi, A., Burke, A., Greenberg-Motamedi, J., and Scott, C. (2008). *Montana Reading First evaluation report 2007–2008.* Portland, OR: Northwest Regional Educational Laboratory.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys.* New York, NY: Wiley.

Russell, S.J. (1997). The role of curriculum in teacher development. In S.N. Friel and G.W. Bright (Eds.), *Reflecting on our work: NSF teacher enhancement in K–6 mathematics*. Lanham, MD: University Press of America.

Russo, A. (2004). School-based coaching: a revolution in professional development—or just the latest fad? *Harvard Education Letter, 20*(4), 1–4.

Schafer, J.L. (1999). NORM: multiple imputation of incomplete multivariate data under a normal model, version 2. Software for Windows 95/98/NT, available from www.stat.psu.edu/~jls/misoftwa.html.

Scheaffer, R.L. (2000). Statistics for a new century. In M. Burke and F.R. Curcio (Eds.), *Learning mathematics for the new century.* Reston, VA: National Council of Teachers of Mathematics.

Scott, C. (with Jennings, J., and Rentner, D.S.). (2006). *Keeping watch on Reading First.* Washington, DC: Center on Education Policy.

Scott, C. (2007). *Moving beyond identification: assisting schools in improvement*. Washington, DC: Center on Education Policy.

Scott, C. (2008). *A call to restructure restructuring: lessons from the No Child Left Behind Act in five states*. Washington, DC: Center on Education Policy.

Smith, T.M., and Ingersoll, R.M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal, 41*(3), 681–714.

Steffy, B.E., and Wolfe, M.P. (1997). *The life cycle of the career teacher: maintaining excellence for a lifetime*. West Lafayette, IN: Kappa Delta Pi.

Steffy, B.E., Wolfe, M.P., Pasch, S.H., and Enz, B.Z. (Eds.). (2000). *Life cycle of the career teacher.* Thousand Oaks, CA.: Corwin Press.

Steiner, L. (2000). *A review of the research literature on scaling up in education: the problem of scaling up in education*. Chicago, IL: North Central Regional Educational Laboratory.

Taylor, J.E. (2005). *Sustainability: examining the survival of schools' comprehensive school reform efforts.* Paper prepared for the annual meeting of the American Educational Research Association, Montreal, Canada.

Thomas, C.S. (2002). *Data analysis knowledge among preservice elementary education teachers.* Paper presented at the International Conference on Teaching Statistics (ICOTS), Cape Town, South Africa. Retrieved October 22, 2007, from www.stat.auckland.ac.nz/~iase/publications/1/10_23_ct.pdf.

U.S. Department of Education, National Center for Educational Statistics, Common Core of Data. (2006). Public Elementary/Secondary School Universe Survey: Year 2003–04 (Final Version 1a). Retrieved August 25, 2006, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006324.

Walpole, S., and McKenna, M.C. (2004). *The literacy coach's handbook: a guide to research-based practice.* New York, NY: Guilford Press.