An Exploration of Instructional Practices that Foster Language Development and Comprehension: Evidence from Prekindergarten through Grade 3 in Title I Schools

July 2017

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

U.S. Department of Education

This page has been left blank for double-sided copying.

An Exploration of Instructional Practices that Foster Language Development and Comprehension: Evidence from Prekindergarten through Grade 3 in Title I Schools

July 2017

Hanley Chiang
Elias Walsh
Mathematica Policy Research

Timothy Shanahan
University of Illinois at Chicago

Claudia Gentile
Alyssa Maccarone
Tiffany Waits
Barbara Carlson
Mathematica Policy Research

Samuel Rikoon
Educational Testing Service

Tracy Rimdzius
Project Officer
Institute of Education Sciences

NCEE 2017-4024
U.S. DEPARTMENT OF EDUCATION

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

U.S. Department of Education
Betsy DeVos
Secretary

Institute of Education Sciences
Thomas W. Brock
Commissioner of the National Center for Education Research
Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance
Ricky Takai
Acting Commissioner

July 2017

The report was prepared for the Institute of Education Sciences under Contract No. ED-04-CO-0112/0011. The project officer is Tracy Rimdzius in the National Center for Education Evaluation and Regional Assistance.

Chiang, H., Walsh, E., Shanahan, T., Gentile, C., Maccarone, A., Waits, T., Carlson, B., & Rikoon, S. (2017). *An Exploration of Instructional Practices that Foster Language Development and Comprehension: Evidence from Prekindergarten through Grade 3 in Title I Schools* (NCEE 2017-4024). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

This report is available on the IES website at http://ies.ed.gov/ncee.

ACKNOWLEDGMENTS

This study benefited from the assistance and contributions of many individuals. In the participating schools, many students, teachers, principals, and prekindergarten directors generously volunteered their time to provide data for the study. Throughout this study, we received helpful input from the study's technical working group and expert consultants, including Thomas Cook, David Dickinson, Barbara Foorman, Chris Lonigan, Paul McDermott, P. David Pearson, Charles Perfetti, Ray Reutzel, Don Rock, Chris Schatschneider, Catherine Snow, and Lihshing Leigh Wang.

The collection and analysis of data for this study required a large effort. Sally Atkins-Burnett, Timothy Bruursema, John Carsley, Melissa Hepburn, Lizabeth Malone, Rene Nutter, Christine Ross, Sara Skidmore, Jillian Stein, and Jerry West provided leadership over important aspects of the collection and processing of the data. Michelle Najarian, Dmitriy Pozynak, Harshini Shah, Jillian Stein, and Yange Xue oversaw key aspects of the analysis. Judy Cannon, Xiaofan Sun, and Clare Wolfendale programmed the data analyses effectively and efficiently. Scott Brimacombe and Sarah Osborn provided excellent project management support. When organizing the study's findings into this report, we received thoughtful input from Jill Constantine. Maura Butler provided editing support, and Kimberly Ruffin handled the production of this report.

This page has been left blank for double-sided copying.

CONTENTS

This page has been left blank for double-sided copying.

I. INTRODUCTION

Reading comprehension—the ability to understand the meaning of text—is a foundational ability that enables children to learn in school and throughout life. Children who struggle with reading comprehension in the third or fourth grade are at high risk for dropping out of school, with detrimental effects on their future employment, income, and participation in the social and political aspects of life (Chrissey 2009).

Unfortunately, many students reach the fourth grade still struggling to comprehend text at a basic level. In a nationally representative sample of fourth graders who were administered the 2015 National Assessment of Educational Progress (NAEP) reading assessment, one-third of the students performed below a basic level (U.S. Department of Education 2015). Moreover, specific groups of children, including children from low-income families and English language learners, are even more likely to struggle to read. On the 2015 NAEP reading assessment, 44 percent of students who qualified for free or reduced-price lunch and 68 percent of English language learners scored below the basic level.

To understand the factors that may contribute to students' struggles with reading comprehension, researchers have explored the relationship between students' early language skill development and their later reading comprehension levels. The language skills considered in such research consist of the range of diverse skills needed to understand and express meaning, including knowledge of phonics, grammar and syntax, vocabulary, and listening comprehension. The evidence indicates that the reading comprehension outcomes of elementary school students are strongly associated with their earlier development of these language skills (Kendeou et al. 2009; National Early Literacy Panel 2008). Therefore, elementary school students who have difficulty reading are very likely to have struggled with early language skills.

Because of the critical importance of early language development and reading comprehension to children's academic achievement, many research studies have focused on identifying instructional practices that support children's language development and promote growth in reading comprehension. Since the mid-1990s, several national panels of experts have synthesized hundreds of small-scale studies (each study with usually fewer than 500 students) on the prevention of reading difficulties and on strategies to teach reading to young children (see the summaries of this research by Snow et al. [1998]; National Institute of Child Health and Human Development [NICHD] [2000]; National Early Literacy Panel [2008]). For example, one prominent panel commissioned by Congress, the National Reading Panel, concluded that explicit teaching of five key language and literacy skills and strategies would lead to higher reading comprehension achievement: (1) phonemic awareness (awareness of the units of sounds that form words), (2) phonics and decoding (knowledge of the relationships between sounds and letter patterns), (3) vocabulary, (4) fluency, and (5) use of comprehension strategies such as summarizing and generating questions about text (NICHD 2000).

However, large public investments to encourage evidence-based instruction in the skill and strategy areas highlighted by the expert panels have not generated significant or meaningful improvements in students' language development and comprehension. In particular, Congress established two major programs—Early Reading First and Reading First—to provide states, districts, and private organizations with grants that supported the adoption of evidence-based

reading instruction by preschools (in Early Reading First) and elementary schools (in Reading First). Although these programs did not specify exactly which practices or interventions to adopt, grantees were expected to use curricula and instructional materials based on scientific evidence, provide professional development to help teachers implement evidence-based instruction, and screen students to identify early reading difficulties. National evaluations of Early Reading First (Jackson et al. 2007) and Reading First (Gamse et al. 2008a, 2008b) found that the programs improved teachers' instructional practices in early literacy and reading. However, the programs had little or no impact on children's language outcomes in preschool (Jackson et al. 2007) or reading comprehension outcomes in the first through third grades (Gamse et al. 2008a, 2008b). In addition, subgroup analyses in the Early Reading First study showed a pattern of negative effects on the language outcomes of children who attended Head Start-funded centers. Well-designed evaluations of other relevant interventions, including more than a dozen preschool curricula (Preschool Curriculum and Evaluation Research Consortium 2008) and enhancements to the Even Start Family Literacy Program focused on research-based literacy instruction (Judkins et al. 2008), found little or no positive effect on children's language development, particularly in prekindergarten.

More recent studies found a similar pattern of inconsistent effects of early childhood interventions on children's language and comprehension skills. These programs were sometimes found to have positive impacts on narrowly defined skills but not on a broad range of language and comprehension outcomes, or to have impacts that did not persist over time. For example, Mashburn et al. (2016) found few positive impacts of a prekindergarten curriculum designed to promote children's development of vocabulary, narrative expression, print knowledge, and phonological awareness and intended to be easily implemented on a large scale. The curriculum led to improvements in students' knowledge of print concepts but did not affect five other language or early literacy outcomes examined by the study. A random assignment evaluation of Tennessee's voluntary prekindergarten program found that children assigned to the intervention performed significantly better than comparison group children on language assessments at the end of the prekindergarten year, but that the comparison group children caught up to the intervention group by the end of kindergarten and surpassed the intervention group on language and reading comprehension assessments in the second and third grades (Lipsey et al. 2015).

Given the modest and inconsistent effects of existing large-scale early literacy interventions, the Institute of Education Sciences at the U.S. Department of Education commissioned this study to investigate additional types of instructional practices that hold potential promise for promoting young children's language development and comprehension. Using an exploratory design, the study team collected extensive information about instructional practices in prekindergarten through grade 3 within Title I schools and examined the relationships between these practices and student growth in a range of language and comprehension outcomes. Findings from this study are intended to help identify potentially promising practices that ought to be studied further and evaluated on a large scale. The study is not designed to make recommendations for classroom practice. As designed, this study makes three key contributions to the existing body of research about the relationships between instructional practices and young children's language and comprehension growth: (1) the exploration of a wide range of instructional practices; (2) the use of student outcome measures that cover a range of language and comprehension skills; and (3) the exploration of the relationship between practices and student growth on a large scale.

First, the study explores a wider range of instructional practices than those that formed the basis for recent federal early literacy programs. As discussed earlier, programs such as Reading First and Early Reading First promoted instructional practices focused on the five skill and strategy areas identified by the National Reading Panel, drawing upon substantial research that had found positive effects of these practices in smaller-scale settings. Because these programs, for the most part, did not have their intended effects when widely implemented, this study collected information on an even broader array of practices to search for those practices that might be related to student growth and could therefore be evaluated further. Beyond the practices emphasized by the expert panels, we considered practices that encourage students' oral language, expose them to knowledge of the world, stimulate higher-order thinking, help them focus on the meaning of texts, and encourage their engagement in the classroom.

Second, the study examines student outcome measures that cover a range of language and comprehension skills. Successful reading comprehension depends on a number of other outcomes that form a foundation for being able to understand text, including a variety of language skills and background knowledge about the social and natural world. For instance, a large synthesis of research found that language measures covering a variety of skills were much more strongly correlated with subsequent reading comprehension than measures focused narrowly on particular skills, such as vocabulary alone (National Early Literacy Panel 2008). However, many previous studies examined the effects of instructional practices on only particular outcomes that were closely aligned with the practices being considered (see Chapter IV for a detailed discussion). For example, although some studies investigated the effects of vocabulary instruction on vocabulary improvement in the early grades (see, for example, Beck and McKeown [2007] and Penno et al. [2002]), almost none determined whether the effects carried over to other aspects of language development and, ultimately, to reading comprehension. There have been exceptions in which a small number of studies (which took place concurrently with the present study) have examined the effects of early-grade language instruction on a range of outcomes and even longer-term comprehension outcomes (see, for example, Lyster et al. [2016] and Dickinson and Porche [2011]), yet most studies have generally examined a smaller set of outcomes. This study contributes to existing research by examining outcome measures that encompass diverse language skills, aspects of background knowledge, and ultimately reading comprehension.

Third, the study examines relationships between practices and student growth on a large scale. The findings are based on data collected from 83 Title I schools in 9 states, in which the study team observed instructional practices in over 1,000 classrooms and administered assessments to nearly 5,000 children in the 2011–2012 school year. The size of this exploratory study is important because its findings are intended to suggest the types of early literacy practices that ought to be evaluated on a large scale.

Given this study's exploratory design, it cannot provide conclusive information about the effectiveness of instructional practices and is not meant to make recommendations for actual classroom instruction. Instead, the goal is to suggest directions for future research on practices that may promote language development and comprehension.

The main body of this report presents, in brief, the study's methods and main findings. Chapter II outlines the design of the study, the types of data collected, and the methods used to

collect and analyze the data. Chapter III presents the findings for all students as well as specific groups of students. Chapter IV discusses the study's contributions to prior research and possible avenues for future research. In the appendices, we provide more detailed information about the selection and characteristics of the study sample (Appendix A); the study instruments and data collection procedures (Appendix B); the analysis methods (Appendix C); additional analyses (Appendix D); and a copy of the classroom observation rubric developed for this study (Appendix E).

II. STUDY DESIGN, DATA, AND METHODS

To examine the relationships between instructional practices and student growth in language and comprehension, we observed instructional practices in a sample of classrooms within Title I schools and administered assessments to students in those classrooms. The following sections provide an overview of the study sample (Section A), how we measured instructional practices (Section B), how we measured students' language development and comprehension (Section C), the approaches we used to analyze the data (Section D), and the study's key limitations (Section E).

A. Study sample

The study sample was designed to maximize variation in student growth across classrooms and schools. This variation was important for identifying instructional practices that were associated with greater student growth.

The study was conducted in 10 districts located in nine states in diverse geographic regions. The districts were purposively selected because they had large numbers of high- and low-performing Title I elementary schools. From each district, we randomly selected samples of high- and low-performing elementary schools that had schoolwide Title I programs and included all grades from prekindergarten through grade 3. Within each school, we randomly selected up to three classrooms in each of these five grades for observation by trained observers. In each of these classrooms, we also randomly selected a sample of students to whom we administered fall and spring assessments.

The final sample for the analysis consisted of 83 schools, 1,035 classrooms, and 4,969 students. Appendix A provides detailed information about the selection and characteristics of the study participants.

B. Measuring instructional practices

Measuring a diverse set of instructional practices that might influence language development and comprehension skills in young children was an important contribution of the study. Existing observation instruments that spanned prekindergarten through grade 3 focused only on general classroom practices. Other instruments that focused on language development and comprehension practices applied to only preschool or only upper elementary and secondary grade levels. The Observation of Language and Literacy Instruction (OLLI) was developed specifically for this study to cover this gap in observation instruments (see Appendix E for a copy of the OLLI, which includes all items, definitions of the key constructs measured, and how they were scored).

To develop the OLLI, the study team conducted a literature review to identify the aspects of instruction that research suggested were related to students' language development and comprehension in listening and reading, resulting in a comprehensive list of instructional practices representing different and sometimes competing theories. The OLLI focused on these diverse aspects of instruction, and the final instrument included 285 items. Most items focused

on six dimensions of language and literacy instruction: (1) teachers' use of language,[1] (2) text-related instruction, (3) vocabulary instruction, (4) teaching of reading comprehension strategies, (5) instruction about world knowledge, and (6) encouragement of higher-order thinking. The OLLI was not designed to measure and compare specific strategies for teaching phonics and oral reading fluency because prior research has studied those strategies extensively.

The sections on text-related and vocabulary instruction included items to capture whether activities occurred as part of pre-reading, during-reading, or post-reading instruction. These distinctions applied regardless of the subject being taught. Specifically, whenever a teacher engaged students in discussing a text that they were about to read (for English language arts, mathematics, social studies, or science), the activity was coded as pre-reading. Whenever a teacher engaged the students in reading a text in class, the activity was coded as occurring during reading; and whenever a teacher engaged students in discussing a text that they had just read (that same day), the activity was coded as post-reading.

In addition to the six dimensions that focused on practices related to language development and literacy instruction, the OLLI included four other dimensions that captured general instructional practices—classroom context, classroom climate, time management, and student engagement. These were developed by borrowing or adapting items from other commonly used observation instruments, including the Classroom Assessment Scoring System (CLASS; Pianta et al. 2006) and Teacher Behavior Rating Scale (TBRS; Landry et al. 2001).

The OLLI included three basic types of items: (1) occurrence, (2) intensity, and (3) quality. Some items recorded the basic occurrence of an action, such as whether or not the teacher talked about the characters in a book. Other items recorded the intensity of an action or amount of a practice, such as how many words were defined during a vocabulary lesson. Still other items focused on the quality of an action, such as the degree to which post-reading discussion was focused on content and was coherent.

In the spring of 2012, the study team recruited and trained observers to conduct classroom observations with the OLLI. Approximately 100 trainees (80 percent with classroom experience as either teachers or teacher's aides, and 100 percent with undergraduate degrees) underwent a 10-day training session, including practice applying each section of the OLLI and two days of practice using the full instrument to rate video recordings of instruction in prekindergarten through grade 3. The training culminated in two days of certification activities in which trainees were required to observe instruction (both video recordings and live instruction) and generate ratings in sufficient agreement with those of the trainers. Approximately 90 percent of the trainees passed the certification criteria by demonstrating 80 percent exact agreement for each dimension of the OLLI.

The trained and certified observers used the OLLI to observe the study classrooms in the spring of 2012. They conducted up to four 90-minute observation sessions per classroom, with each of the four sessions conducted by a different observer on a different day. Typically, two of

[1] Note that, through the OLLI, we captured information about the ways in which teachers encouraged student language and how often they did so, how often they spoke with students, and the quality (clarity and correctness) of their language, but did not capture information about students' generating language during instruction.

the sessions occurred in the morning (when English language arts instruction was most likely to take place) and two occurred in the afternoon (when social studies or science instruction was most likely to take place) so that we would observe a range of practices in each classroom. Each observation session consisted of six 15-minute segments, for a total of 90 minutes of observed instruction per session. After each 15-minute segment, observers spent 5 additional minutes assigning scores for that segment on all OLLI items. On each item, the score for an observation session was the average score across the six segments.

Although the number of items on the OLLI (285) was large, approximately two-thirds of the items were coded only when specific instructional activities were occurring. For any given observation segment, observers coded the subsample of items that corresponded to the types of activities they were witnessing. For example, 124 items on teachers' approaches to teaching texts were applicable only if any text-related instruction was observed, which occurred in only about one-third of observation segments in prekindergarten and kindergarten and one-half of segments in grades 1 through 3 (Appendix D, Table D.10). Observers skipped items that were not applicable to a given observation segment.

Despite having to assign scores on a large number of items, observers demonstrated consistency with each other in how they assigned scores. Although each observation session usually had only one observer, the study assigned multiple observers to some of the observation sessions to check for consistency across observers.[2] On average across all items, observers who rated the same segment of instruction assigned exactly the same score 83 percent of the time. Appendix B provides more information on the development of the OLLI, the observation procedures, and consistency across observers.

C. Measuring student growth

A key contribution of this study, as discussed in Chapter I, was to examine the relationships between instructional practices and students' growth in a range of language and comprehension outcomes. We administered assessments to participating students in both fall 2011 and spring 2012 to measure their growth in several domains of language and comprehension (Table II.1). As we described in Chapter I, although reading comprehension is the ultimate outcome of interest, researchers have identified a variety of related outcomes that form a foundation for successful reading comprehension. Therefore, in addition to measuring reading comprehension directly, this study also included measures of three other outcomes that research has identified to be important for reading comprehension: (1) basic language skills, a multidimensional outcome captured by a composite measure; (2) listening comprehension, the aspect of language development that is conceptually closest to reading comprehension; and (3) background knowledge in science and social studies, a key input into comprehension.

We measured each of the four outcomes with a different assessment measure (Table II.1). Each measure had evidence of reliability and validity, was appropriate for measuring student growth over a school year, and had been used in prior research studies (Wiig et al. 2004; Semel

[2] Every observer co-observed one full observation session (typically covering six 15-minute segments of instruction) with at least one other observer. A total of 42 observation sessions, encompassing more than 200 observation segments, were included in this assessment of consistency across observers.

et al. 2003; U.S. Department of Education 2002, 2004; Woodcock et al. 2001, 2007; Pollack et al. 2005). Assessments to measure basic language skills and listening comprehension were administered to students in all five grades, from prekindergarten through grade 3. The assessment of background knowledge was administered to students in prekindergarten through grade 1, and the assessment of reading comprehension was administered to students in grades 2 and 3. In what follows, we provide an overview of these outcomes and how we measured them. Appendix B provides a detailed explanation of the outcome measures used in the study.

Table II.1. Student assessments administered in the study

Domain of language and comprehension	Name of assessment	Grades
Basic language skills	Clinical Evaluation of Language Fundamentals[a]	PK–3
Background knowledge	Early Childhood Longitudinal Study–Kindergarten Class of 1998–99 General Knowledge Assessment[b]	PK–1
Listening comprehension	Woodcock-Johnson III Tests of Achievement, Oral Comprehension Subtest[c]	PK–3
Reading comprehension	Early Childhood Longitudinal Study–Kindergarten Class of 1998–99 Third Grade Reading Assessment[d]	2–3

Source: Authors' compilation.

[a] We administered two Clinical Evaluation of Language Fundamentals (CELF) assessment batteries: the CELF Preschool – Second Edition (Wiig et al. 2004) for prekindergarten and kindergarten students, and the CELF – Fourth Edition (Semel et al. 2003) for students in 1st, 2nd, and 3rd grades.

[b] U.S. Department of Education (2002).

[c] Woodcock et al. (2001, 2007).

[d] U.S. Department of Education (2004); Pollack et al. (2005). Additional items appropriate for grade 2 come from the Early Childhood Longitudinal Study–Kindergarten Class of 2010–11 Second Grade Reading Assessment (Tourangeau et al. 2017).

PK = prekindergarten.

Basic language skills. As discussed in Chapter I, basic language skills encompass multiple types of skills needed to understand and express meaning. The language assessment we administered in all grades, the Clinical Evaluation of Language Fundamentals (CELF; Semel et al. 2003; Wiig et al. 2004), covered a range of skills that research has found to be important precursors to reading comprehension, including students' receptive vocabulary knowledge (understanding words when spoken) and expressive vocabulary knowledge (expressing thoughts orally); their understanding of syntax; and their understanding of different units of meaning within words, such as prefixes and suffixes. We chose a composite measure of language skills because such measures are much more strongly correlated with subsequent reading comprehension than are measures of only one skill alone, such as vocabulary (National Early Literacy Panel 2008).

Background knowledge. Students' background knowledge includes their familiarity with basic concepts (such as space and time) and with the social, physical, and biological world. Researchers have suggested that background knowledge helps students extract meaning from texts (Hirsch 2003, 2006; Hoover and Gough 1990). For example, an understanding of time enables students to sequence events in a story. Background knowledge can also help students understand the context of the words they read, beyond simply understanding the words' literal definitions (Snow et al. 1998). In prior research, background knowledge of social studies and

science content in kindergarten has been positively associated with reading achievement in grades 1, 3, and 5 (Claessens et al. 2009; Duncan et al. 2007).

We assessed the background knowledge of students in prekindergarten, kindergarten, and grade 1 with the Early Childhood Longitudinal Study–Kindergarten Class of 1998–99 (ECLS–K) General Knowledge Assessment (U.S. Department of Education 2002). This measure included assessment of both science (including earth and space, life, and physical sciences) and social studies (including culture, history, geography, government, and economics).

Listening comprehension. Listening comprehension, the ability to understand spoken language, is the aspect of language development that is conceptually closest to reading comprehension, the ability to understand written language. Evidence also supports a close connection between these two outcomes. A meta-analysis of 30 independent studies indicates a relationship between kindergarteners' listening comprehension and their later reading comprehension through age 7 (National Early Literacy Panel 2008).

Because our composite measure of language skills covered a variety of skills, we chose also to administer an assessment specifically capturing listening comprehension, given its close connection to reading comprehension. We assessed listening comprehension with the Woodcock-Johnson III (W-J III) Tests of Achievement, Oral Comprehension subtest (Woodcock et al. 2001, 2007) in all of the study grades from prekindergarten through grade 3. The assessment asked students to verbally supply the missing key word that completed an oral passage.

Reading comprehension. Although reading comprehension is the ultimate outcome of interest, it generally cannot be measured with sufficient validity before grade 2. Scores on reading comprehension assessments in grade 1 and earlier are often misleading, as they are aligned too closely with decoding or word reading skills to represent a truly independent measure of reading comprehension (Francis et al. 2005; Keenan et al. 2008; Nation and Snowling 1997). For this reason, we measured reading comprehension only in grades 2 and 3 and focused on the foundations of reading comprehension in grade 1 and below, as described earlier.

We measured reading comprehension with the ECLS-K Third Grade Reading Assessment, augmented with items that were also appropriate for second-grade students. This assessment measured students' ability to identify the main point of a written passage, interpret the passage, connect it to their background knowledge, and evaluate its key features.

Although, on average, students in the study performed below the national average on each assessment, their performance differed in a manner that the assessments could reliably capture. In fact, on all of the assessments, test scores differed substantially across students. For example, the top one-fifth of students in the study typically scored above the 50th percentile nationwide, whereas the bottom one-fifth of students in the study typically scored below the 20th percentile nationwide (Appendix A, Table A.6). These differences in test scores were also highly reliable, with no more than one-tenth of the variation in scores being attributable to measurement errors that were not indicative of true skills (Appendix B, Table B.9).

Conceptually, each of these four outcomes had the potential to be more strongly related to certain types of instructional practices covered by the OLLI than to others. For example, students' growth on the background knowledge assessment could be especially related to the dimension of the OLLI on world knowledge instruction, given that this dimension covered aspects of teaching aimed at improving students' knowledge of facts and concepts. Likewise, students' growth in reading comprehension could be particularly associated with the OLLI dimension on the teaching of reading comprehension strategies. Despite the close conceptual connection between specific outcomes and particular dimensions of instructional practices, we followed an exploratory approach, described in the next section, of considering potential relationships between any of the four outcomes and any of the practices measured in our study.[3]

D. Analytic approach

The study's analysis was aimed at identifying instructional practices that were associated with student growth in language and comprehension over one school year. The key steps in the analysis consisted of (1) creating summary measures of instructional practices, (2) measuring teachers' contributions to student growth, and (3) assessing the relationships between the summary measures of practices and teachers' contributions to student growth. Appendix C provides technical details on each of these steps.

1. Creating summary measures of instructional practices

The 285 items on the OLLI captured many specific aspects of instruction. Examining the relationship between each of these items and student growth would have led to many imprecisely estimated relationships. This would make it difficult to extract clear hypotheses on the most promising ways to promote language and comprehension growth. To sharpen the study's focus on a smaller number of instructional practices, we used data-driven approaches to identify groups of items that were strongly related to each other because they reflected the same underlying instructional practice. Each group of items formed a summary measure of an instructional practice that could be examined in subsequent analyses of relationships with student growth. Below, we briefly describe four key steps to create the summary measures (Figure II.1).

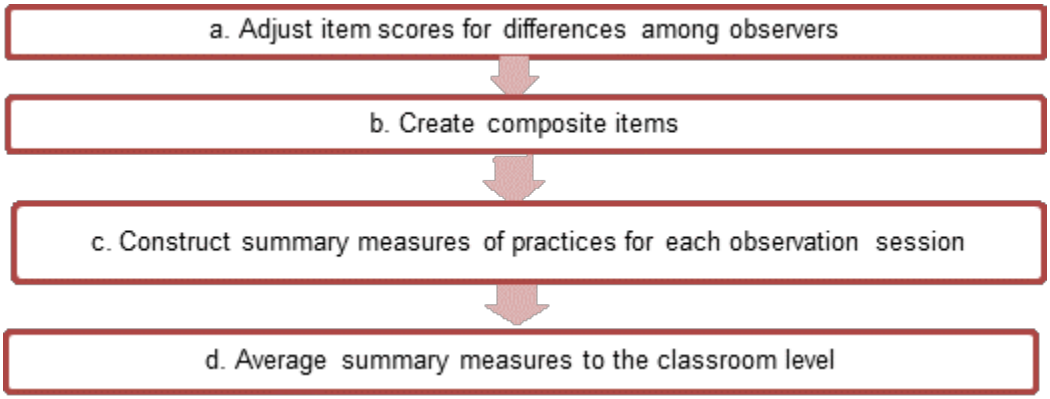
a. Adjust item scores for differences among observers. For each OLLI item, we identified systematic differences in the scores given by different observers. Those differences, known as observer effects, could generate differences in item scores across classrooms that did not reflect true differences in practices. To address this problem, we removed those differences from the item scores.

b. Create composite items. The main objective in the analysis of the OLLI data was to identify a smaller number of well-defined instructional practices that underlay the large number of OLLI items. However, standard techniques to identify underlying behaviors from observed

[3] Given that reading comprehension is the ultimate outcome of interest for which the other three outcomes provide a foundation, researchers may like to know the extent to which each instructional practice affects reading comprehension via its effect on the other three outcomes. Such an analysis, known as a mediator analysis, would be potentially useful for studies designed to measure the effects of these practices. Because this study was meant only to suggest practices for further research—not to measure effects in a conclusive way—we did not conduct a mediator analysis.

items, such as exploratory factor analysis, could not have incorporated such a large number of OLLI items—285 in total. With such a large number of items, the behaviors identified by a factor analysis would be expected to fit the data poorly (Marsh et al. 2014). For this reason, we first reduced the number of items by constructing various composite items before attempting to identify well-defined instructional practices.

Figure II.1. Creating summary measures of instructional practices



We combined some items into composite items if they pertained to the same aspect of instruction and were already listed together under the same prompt in the OLLI. For example, a list of items that shared the prompt, "What techniques did the teacher use to help students expand their use of language?" captured several techniques that had the common objective of expanding students' language use. These items were well suited to be combined into composite items about the frequency and diversity of techniques to help expand students' use of language.

Within each list of items, we used a data-driven technique known as principal components analysis to determine how many composite items would be created and which items would be combined together to form those composite items. Principal components analysis linearly combined the original items into composites that retained as much of the variation in scores on the original items as possible. Because this stage of the analysis was not meant to create the final summary measures of instructional practices, it was important to retain as much of the original variation in item scores as possible so that this information could be used in subsequent stages of analysis to create the final measures of instructional practices. For this reason, principal components analysis was particularly suited to creating composite items. In total, we created 12 composite items from four large lists of original items. Appendix C, Tables C.1 to C.12 show the composite items and the original items from which they were constructed.

This process resulted in a reduced set of 89 items—12 composite items, plus 77 original items that were not incorporated into composites. On the one hand, 89 items, if analyzed individually for relationships with student growth, would still yield a large number of imprecisely estimated relationships with few clear lessons. On the other hand, these items were sufficiently reduced in number to permit standard techniques to identify coherent groups of items representing the same underlying instructional practice. We describe next the process for identifying the underlying practices that served as the focus of the remainder of the study.

c. Construct summary measures of practices for each observation session. We used exploratory factor analysis to identify groups of items (both composite and individual items) that were highly correlated with each other because they reflected a common "factor"—a single underlying instructional practice. Each group of items gave rise to a summary measure of an instructional practice, producing a score on that practice for every observation session.[4]

The exploratory factor analysis identified a diverse set of 13 instructional practices (Table II.2). Appendix C, Tables C.13 to C.25 show all of the items that contributed to the summary measures of those 13 practices. Some items—including several items about engaging students in defining new words and helping students use comprehension strategies—measured actions that were rarely observed (in fewer than 10 percent of observation segments), implying that the summary measures that included those items had limited variation. Due to the limited variation in some practices, the study's standard for assessing whether practices were significantly associated with student growth was less stringent than conventional standards, as discussed later in this chapter.[5]

As noted before, the items from the OLLI that were used to develop the summary measures captured the occurrence, intensity, or quality of specific aspects of instruction. For an observation session to be assigned a high score on a particular summary measure, the teacher had to have engaged in the targeted practice in many observation segments (frequent occurrence), repeatedly within each segment (high intensity), and in a manner thought to be desirable (high quality). For example, for the summary measure of encouraging students' oral language, observation sessions with high scores were those in which the teacher (1) engaged in multiple practices in multiple segments to expand students' use of language (high scores on the occurrence items); (2) spoke with students most of the time within each segment (high scores on the intensity items); and (3) used clear and correct language, and focused the talk on instruction or content, not directions or behavior management (high scores on the quality items).

The methods that we used to generate the 13 summary measures shown in Table II.2 reflect the aims of this study only. Specifically, the principal components and exploratory factor analysis methods used for this study (described in more detail in Appendix C) were selected with the sole objective of reducing the 285 OLLI items to a smaller set of interpretable measures that could account for the correlations among OLLI items in our study sample. Given that this was not a measurement study, the goal was not to establish instructional practice scales that could be used in future research or applied settings. Therefore, the 13 summary measures in Table II.2 would not necessarily explain correlations among items if the OLLI were used to measure instruction in a different sample of classrooms.

[4] See Appendix C for an explanation of why exploratory factor analysis was better suited for creating the final summary measures than principal components analysis.

[5] No summary measures of instructional practices were composed of rarely observed items alone. Also, as mentioned earlier, some observation sessions had multiple observers. Appendix C, Table C.26 provides information on the degree of consistency between instructional practice scores from different observers who rated the same session.

Table II.2. Instructional practices identified from the exploratory factor analysis

Name of practice	Examples of observation items that reflect the practice	Total number of contributing items	Internal consistency of scores on the practice
1. Encouraging students' oral language	• Teacher asked open-ended questions • Teacher allowed students sufficient time to respond to questions	11	0.84
2. Focusing on phonics and grammar during reading	• During reading, teacher discussed grammar, mechanics, or spelling • During reading, teacher discussed letters or words (sounding out letters or words, rhyming words, word recognition)	6	0.77
3. Engaging students in defining new words during pre-reading	• Teacher or students used more than one approach to define a word during pre-reading • Extent of students' involvement in defining words during pre-reading	5	0.91
4. Engaging students in defining new words during reading	• Teacher or students used more than one approach to define a word during reading • Extent of students' involvement in defining words during reading	5	0.91
5. Engaging students in defining new words during post-reading	• Teacher or students used more than one approach to define a word during post-reading • Extent of students' involvement in defining words during post-reading	5	0.90
6. Engaging students in defining new words outside of reading	• Outside of reading, teacher or students defined words by providing additional descriptors • Outside of reading, students had some involvement in defining words	7	0.78
7. Focusing on the meaning of texts during pre-reading	• Extent to which teacher organized talk about the content of a text during pre-reading • Extent of detail that teacher used to talk about the content of a text during pre-reading	3	0.94
8. Focusing on the meaning of texts during reading	• Extent to which teacher organized talk about the content of a text during reading • Extent of detail that teacher used to talk about the content of a text during reading	4	0.91
9. Focusing on the meaning of texts during post-reading	• Extent to which teacher organized talk about the content of a text during post-reading • Extent of detail that teacher used to talk about the content of a text during post-reading	3	0.93
10. Helping students make connections between their prior knowledge and texts	• Teacher connected big ideas in a text to students' prior knowledge • Teacher connected specific details in a text to students' prior knowledge	8	0.72
11. Teaching students to use other comprehension strategies	• Specificity of teacher's explanation of how to use a comprehension strategy • Extent to which teacher explained why a comprehension strategy should be used	6	0.95
12. Focusing on world knowledge	• Time spent in teaching information/facts about the social or natural world • Number of pieces of information about the world taught	11	0.91
13. Focusing on higher-order thinking	• Time spent in encouraging students to use higher-order thinking • Number of questions that asked students to explain their answers or thinking	4	0.86

Table II.2. *(continued)*

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Contributing items are those that had a factor loading of at least 0.30—that is, a one standard deviation increase in the underlying instructional practice was associated with at least a 0.30 standard deviation increase in scores on the item. In the second column, examples of observation items that reflect the practice include some items that contributed to a composite item that, in turn, contributed to the summary measure of the instructional practice. Internal consistency is measured by Cronbach's alpha.

d. Average summary measures to the classroom level. The goal was to accurately measure a teacher's usual practices in his or her classroom—not just his or her practices from a single observation session. As discussed earlier, each classroom had up to four observation sessions. Accordingly, scores on the summary measures were averaged across the observation sessions in each classroom to generate the final measures of each classroom's instructional practices.

When measuring the average practice in each classroom, an important challenge was that differences in observed practices between classrooms did not always represent real differences in teachers' usual practices. Some differences in observed practices may, instead, have been due to the time of day in which observations occurred, or chance events that caused the teachers' performance during the observations to be better or worse than their usual instruction. Removing variation that did not reflect real differences in teachers' usual practices could increase the likelihood of identifying practices that were related to student growth. Accordingly, we adjusted the measures in the following two ways:

- **Accounting for the time of day in which observations occurred.** Within the same classrooms, afternoon sessions had lower scores on instructional practice measures than morning sessions, potentially because of teacher and student fatigue or different subject matter taught. Although most classrooms (68 percent) had equal numbers of morning and afternoon sessions, some (26 percent) had more morning than afternoon sessions, and others (6 percent) had more afternoon than morning sessions. We adjusted scores on the summary measures of practices so that classrooms would not have systematically better or worse scores simply because they were observed in more morning or afternoon sessions.
- **Accounting for the limited number of observations per classroom.** Because a teacher's practices could vary from one lesson to the next, the four or fewer observations conducted in each classroom could, by chance, have missed the fuller picture of the teacher's usual instructional practices. Using a technique known as empirical Bayes shrinkage, we estimated how much variation in the summary measures came from measurement error due to limited numbers of observations per classroom, and we filtered out this variation.[6]

2. Measuring teachers' contributions to student growth

When measuring students' language and comprehension outcomes, this study focused on the growth that students made from the fall to the spring of the study school year. The advantage of examining growth, rather than just end-of-year performance, was that it took into account the

[6] For each of the 13 measures of instructional practices, Appendix C, Table C.27 shows the reliability of the classroom-level scores—that is, the percentage of the variation in those scores that represented real differences in teachers' usual practices rather than measurement error—before such measurement error was filtered out.

skills that students had before being taught by their current teachers. Of all the differences in growth across students, the study focused on only the differences due to teachers' contributions—rather than, for example, the students' demographic background—because those were the differences that could result from the teachers' instructional practices.

Specifically, teachers' contributions to student growth were calculated in two steps. First, we estimated a regression model to predict each student's test scores in the spring based on his or her (1) fall score on the same assessment, (2) fall scores on the other administered assessments, and (3) background characteristics. Next, we calculated each teacher's contribution to student growth as the average difference between the actual and predicted spring test scores of his or her students. We used separate models for each grade span (prekindergarten and kindergarten in the lower grade span, and grades 1 to 3 in the upper grade span) and assessment.

3. Assessing the relationships between instructional practices and teachers' contributions to student growth

To assess the relationships between instructional practices and teachers' contributions to student growth, we estimated a final set of regression models. In each model, the estimates of teachers' contributions to student growth constituted the dependent variable, and a summary measure of an instructional practice was the main independent variable. We estimated separate models for each grade span and assessment. Estimating separate models for each grade span accounted for the possibility that instructional practices could have different relationships with language development and comprehension for younger and older children. Each model in the main analysis included a single instructional practice measure, and supplemental analyses included all 13 summary measures simultaneously.

As discussed in Chapter I, the ultimate purpose of assessing relationships between practices and growth was to identify practices that may be worth further study, which we call *potentially promising* practices. In this report, a practice is considered potentially promising in a specific grade span if it had a positive, statistically significant relationship with student growth on at least one outcome and no significant negative relationships with any other outcome in that grade span. Tests of statistical significance used a level of 0.10.

Given the disappointing results of previous large-scale evaluations of scientifically-based reading interventions, we sought to identify as many practices as possible for which initial evidence could warrant further study. Some practices that are worth further evaluation may not have been associated with student growth at a 0.05 significance level, or may not have been associated with growth on multiple outcomes, due to limited variation in the practice across the study classrooms. Setting lenient criteria for being considered potentially promising helped to avoid overlooking these types of practices. At the same time, because we examined many relationships between practices and growth and set lenient criteria for identifying potentially promising practices, some practices may have been incorrectly identified just by chance. In supplemental analyses (see Appendix D), we used more stringent significance levels for statistical tests and applied corrections that accounted for having examined a large number of relationships.

As noted earlier, we identified potentially promising practices separately in the lower grades (prekindergarten and kindergarten) and upper grades (grades 1 through 3). By doing so, we

accounted for the possibility that younger and older children may benefit from different sets of practices. At an early stage of the study, we decided to examine relationships separately by grade span because our preliminary analyses found that the prevalence of most of the practices examined in this study differed between the lower and upper grades. For example, upper-grade teachers focused more heavily on nearly all text-related practices and higher-order thinking than lower-grade teachers did; teachers in the lower grades put greater emphasis on encouraging students' oral language and promoting world knowledge. These differences suggest that, at the very least, the teachers themselves may have believed that certain practices were more effective for younger or older children. Because practices might have different relationships with student growth at different ages, we did not require practices to be related to growth in both grade spans to be considered potentially promising.

E. Limitations

Although this study contributes to our understanding of the relationships between instructional practices and young children's language development and comprehension, there are some limitations to the findings. First, the relationships that the study identified between instructional practices and student growth do not demonstrate that teacher practices caused changes in student outcomes. For example, teachers who use one practice may also tend to combine it with other practices that our study did not measure. If so, the relationships that we found could partly reflect the effects of those other, unmeasured practices. Accordingly, the purpose of the study was to identify practices that might be worth evaluating further to determine their ultimate effectiveness—not to propose the implementation of these practices in classrooms.

Second, although classrooms differed considerably in student growth (as intended by the study design), we could identify a significant relationship between an instructional practice and student growth only when there was also meaningful variation across classrooms in the instructional practice. As discussed earlier, some practices had less variation than others. The lack of significant results for a practice could be due either to the lack of a real relationship with growth or to our not having observed enough classrooms with high and low scores on the practice.

Third, the study measured classroom practices in a limited portion of the school year. As discussed earlier, a key strength of the study was that each teacher's practices were observed in multiple, lengthy sessions conducted by different observers; all of these study design features improved the accuracy with which the study measured teachers' practices. However, these observation sessions occurred within a three-week period in the spring of 2012. Therefore, our measures of practices accurately captured instructional quality at a particular point in the spring term, but not necessarily instructional quality over the whole school year. To the extent that the study's measures of practices were not representative of what teachers did over the whole year, we may find fewer relationships between practices and student growth than we would otherwise have found with yearlong measures of practices.

Fourth, although the classroom observations captured a wide range of practices, they were not designed to capture certain aspects of instruction that might influence student growth. Given that each observation session encompassed 90 minutes of instruction, observers could not determine whether teachers connected instruction coherently across a whole school day—for

instance, by linking multiple lessons in the day repeatedly back to a new concept or term. The observation instrument was also not designed to measure the frequency with which teachers changed classroom structures—a common approach for enhancing engagement—nor could it capture students' level of exposure to challenging subject matter and high-quality instructional materials. Observers did not have access to teachers' lesson plans, so they could not assess whether the observed instruction was planned or spontaneous. Moreover, as with any classroom observation instrument, this study's instrument was not suited to capture details about the quality of teachers' language, which could be more accurately measured through audio or video recordings of lessons.

Fifth, the participating districts were not a representative sample of all U.S. districts with Title I schools. Because we chose districts with large numbers of high- and low-performing Title I schools, the participating districts were larger than the average district with Title I schools and tended to have more variation in student performance across schools. Likewise, the participating schools were not a representative sample of all Title I schools. The participating schools had higher concentrations of low-income students and racial and ethnic minorities—81 percent of students in the participating schools received free or reduced-price lunch, and 94 percent were nonwhite (Appendix A, Table A.4). Moreover, because schools were selected for the study on the basis of being consistently high- or low-performing, the set of participating schools was more polarized in student performance than the national population of Title I schools. Results from this study may not necessarily apply to districts, schools, and students with substantially different characteristics.

This page has been left blank for double-sided copying.

III. RELATIONSHIPS BETWEEN INSTRUCTIONAL PRACTICES AND TEACHERS' CONTRIBUTIONS TO STUDENT GROWTH IN LANGUAGE AND COMPREHENSION

In this chapter, we report findings on the relationships between 13 instructional practices and student growth. In each grade span, we highlight practices that were positively related to student growth on at least one outcome measure and were not negatively related to growth on any other outcome in that grade span. As discussed in Chapter II, we consider those practices to be potentially promising—that is, worth further study.

Practices can be identified as potentially promising in the lower grade span (prekindergarten and kindergarten), the upper grade span (grades 1 to 3), or both. The analysis was based on test scores and practices in 378 classrooms in the lower grades and 657 classrooms in the upper grades. Because we assessed reading comprehension only in grades 2 and 3, results for this outcome pertained to the upper grades only and exclude grade 1 classrooms. Also, we assessed background knowledge only in prekindergarten through grade 1, so our analysis of this outcome in the upper grades included only grade 1 classrooms.

A. Main results

In each grade span, we found several practices to be potentially promising. Some were related to the growth of both students in prekindergarten and kindergarten and students in grades 1 to 3. Two practices—helping students make connections between their prior knowledge and the texts they read, and focusing on higher-order thinking—are potentially promising in both the lower grades (Table III.1) and the upper grades (Table III.2). Others were related to student growth in only one grade span. Engaging students in defining new words during reading, focusing on the meaning of texts during pre-reading, and focusing on world knowledge are potentially promising practices in the lower grades only. Encouraging students' oral language, engaging students in defining new words during post-reading, and teaching students to use other comprehension strategies are potentially promising practices in the upper grades only.

When we identified the student outcome measures to which these potentially promising practices were related, we found that most of these practices were related to only one outcome measure. In the lower grades, each of the five potentially promising practices was positively related to only one student outcome, with nearly all (four out of five) of those practices being related to basic language skills growth (Table III.1). In the upper grades, only one practice—helping students make connections between their prior knowledge and texts—was positively related to multiple student outcomes (Table III.2). The remaining potentially promising practices in the upper grades were each associated with only one outcome, which varied across practices.

In addition, the relationships between instructional practices and student growth in language and comprehension were generally small. No individual practice explained more than 14 percent of the variation in growth across classrooms on any student outcome (Appendix D, Tables D.1 through D.5).

Although practices had inconsistent and generally small relationships with student growth, the evidence indicates that most of the positive relationships we found were more than just

statistical flukes. In total, we found 13 positive and statistically significant relationships between instructional practices and student growth out of the 88 tested. This is more than might have happened by chance, which suggests some practices are truly associated with student learning.

Table III.1. Relationships between instructional practices and student growth in language and comprehension in prekindergarten and kindergarten

Instructional practice	Student outcome			Promising practice?
	Basic language skills	Background knowledge	Listening comprehension	
1. Encouraging students' oral language				No
2. Focusing on phonics and grammar during reading			-	No
3. Engaging students in defining new words during pre-reading		-	-	No
4. Engaging students in defining new words during reading	+			Yes
5. Engaging students in defining new words during post-reading				NA[a]
6. Engaging students in defining new words outside of reading				No
7. Focusing on the meaning of texts during pre-reading	+			Yes
8. Focusing on the meaning of texts during reading		-		No
9. Focusing on the meaning of texts during post-reading				No
10. Helping students make connections between their prior knowledge and texts			+	Yes
11. Teaching students to use other comprehension strategies				No
12. Focusing on world knowledge	+			Yes
13. Focusing on higher-order thinking	+			Yes

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The table includes data from 378 prekindergarten and kindergarten classrooms.

A practice is considered potentially promising if there was at least one positive and significant relationship and no negative and significant relationships.

+ Positive and significantly different from zero at the .10 level, two-tailed test.

- Negative and significantly different from zero at the .10 level, two-tailed test.

[a] Within the lower grades, we did not find evidence that teachers in the study differed in the usual extent to which they engaged students in defining new words during post-reading. Therefore, we did not examine the relationship between this practice and student growth in the lower grades.

NA = not applicable.

Table III.2. Relationships between instructional practices and student growth in language and comprehension in grades 1 to 3

Instructional practice	Student outcome				Promising practice?
	Basic language skills	Background knowledge	Listening comp.	Reading comp.	
1. Encouraging students' oral language		+			Yes
2. Focusing on phonics and grammar during reading					No
3. Engaging students in defining new words during pre-reading					No
4. Engaging students in defining new words during reading		+		-	No
5. Engaging students in defining new words during post-reading				+	Yes
6. Engaging students in defining new words outside of reading					No
7. Focusing on the meaning of texts during pre-reading					No
8. Focusing on the meaning of texts during reading	-	+			No
9. Focusing on the meaning of texts during post-reading					No
10. Helping students make connections between their prior knowledge and texts		+		+	Yes
11. Teaching students to use other comprehension strategies			+		Yes
12. Focusing on world knowledge	-				No
13. Focusing on higher-order thinking		+			Yes

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

A practice is considered potentially promising if there was at least one positive and significant relationship and no negative and significant relationships.

+ Positive and significantly different from zero at the .10 level, two-tailed test.

- Negative and significantly different from zero at the .10 level, two-tailed test.

The remainder of this section discusses the findings for each potentially promising practice in more detail.

Encouraging students' oral language (practice 1)

Some teachers put more emphasis on encouraging students' oral language than did other teachers. Teachers who encouraged students' oral language tended to spend large amounts of instructional time talking with students. Their language was clear and correct, and they used different methods for encouraging students to use language, such as reminding students to use complete sentences, asking students open-ended questions, and allowing students sufficient time to respond to those questions.

We found some evidence that encouraging students' oral language is a potentially promising practice. It was associated with more growth in background knowledge in grade 1 (Table III.2).

Engaging students in defining new words during reading (practice 4) and post-reading (practice 5)

If engaging students in defining new words enables them to acquire, retain, and use a larger vocabulary, it may contribute to enhanced language and comprehension skills. Teachers in the study were observed engaging students in defining new words before, during, and after reading instruction as well as outside of reading instruction. We found mixed evidence on whether practices that engage students in defining new words are potentially promising.

Engaging students in defining new words *during reading instruction* was associated with more growth in basic language skills in prekindergarten and kindergarten (Table III.1), and more background knowledge growth in grade 1 (Table III.2). Also, engaging students in defining new words *during post-reading instruction* was associated with more reading comprehension growth in grades 2 to 3 (Table III.2).

However, engaging students in defining new words *during reading instruction* was also associated with less reading comprehension growth in grades 2 to 3 (Table III.2). Therefore, this practice is potentially promising in the lower grades but not the upper grades. We found no other potentially promising vocabulary-related practices.

Focusing on the meaning of texts during pre-reading (practice 7)

When teaching a text, some teachers chose to focus more heavily on the meaning of the text—for instance, by talking about the important information or the plot of a story—rather than other reading readiness skills, such as phonics or grammar. Instruction about the meaning of a text can occur before, during, or after the teacher or students read the text.

Of the three practices that focused on the meaning of texts, only one—focusing on the meaning of texts during *pre-reading* instruction—is a potentially promising practice. It was associated with more growth in basic language skills in the lower grades (Table III.1).

Making prior knowledge connections (practice 10) and teaching other comprehension strategies (practice 11)

To varying degrees, teachers in the study used two types of practices for intentionally enhancing students' comprehension of texts. One practice involved helping students make connections between their prior knowledge (the knowledge that they bring to a text) and the texts they read. Another practice was to teach students to use other reading comprehension strategies, such as predicting, summarizing, and questioning.

Helping students make connections between their prior knowledge and texts and teaching students to use other comprehension strategies are both potentially promising practices. Helping students make connections between their prior knowledge and texts was the only practice associated with growth in two different student outcomes within the same grade span. In the upper grades (Table III.2), making prior knowledge connections was associated with more growth in background knowledge (grade 1) and reading comprehension (grades 2 to 3). It was

also associated with more listening comprehension growth in prekindergarten and kindergarten (Table III.1). Teaching students to use other comprehension strategies was associated with more listening comprehension growth in grades 1 to 3 (Table III.2).

Focusing on world knowledge (practice 12)

Increasing students' knowledge of the world may indirectly promote their language and comprehension growth—for instance, by enhancing students' understanding of concepts and the contexts for the texts they read. Focusing on world knowledge could include teaching general knowledge about aspects of daily life and subject-specific facts and concepts.

Focusing on world knowledge is a potentially promising practice in the lower grades but not the upper grades. The practice was associated with more growth in basic language skills in the lower grades (Table III.1) but less growth in basic language skills in the upper grades (Table III.2).

Focusing on higher-order thinking (practice 13)

Teachers can encourage higher-order thinking by asking questions and/or engaging students in tasks that require students to analyze, evaluate, or synthesize information, apply their knowledge to new situations, explain their thinking, and develop new ideas. To the extent that encouraging higher-order thinking enhances students' ability to process and use information, it may increase their language and comprehension skills.

Focusing on higher-order thinking is a potentially promising practice in both the lower and upper grades. This practice was associated with more growth in basic language skills in prekindergarten and kindergarten (Table III.1). It was also associated with more background knowledge growth in grade 1 (Table III.2).

B. Alternative approaches to measuring relationships between practices and growth

To assess whether the main results were sensitive to the methods used for measuring relationships between instructional practices and student growth, we explored a variety of alternative approaches to measuring these relationships. We then determined whether the potentially promising practices reported in Tables III.1 and III.2 continued to be identified as potentially promising under the following alternative approaches:

- **Accounting for other practices that teachers may have used:** In the main findings, some portion of the relationship between a practice and student growth could reflect the influence of other practices the teachers used. In an alternative approach, we accounted for the extent to which teachers used the 12 other instructional practices examined by this study when measuring the relationship between each practice and student growth. This represents a more conservative way of exploring the relationships between instructional practices and student growth.
- **Accounting for prerequisite actions:** Certain instructional practices could occur only if teachers performed a specific prerequisite action. For example, teachers could engage students in defining new words during reading only if they were already reading a text to

students or guiding students' reading of a text. In the main approach reported above, comparisons between classrooms with low and high scores on an instructional practice could partially reflect differences in the frequency of the prerequisite action. Instead, researchers may be interested in how well a practice promotes student growth in situations where the prerequisite action has already occurred—for example, how well vocabulary instruction during reading promotes student growth given that a teacher has already decided to do a reading lesson. To address this question, we explored an alternative approach that accounted for the frequency of prerequisite actions when measuring relationships between instructional practices and student growth.

- **Using a more stringent significance level for statistical tests:** The main approach reported above identified a practice as being positively related to a student outcome if its relationship with the outcome was positive and statistically significant at the 10 percent level. To reduce the likelihood of finding relationships that occurred simply by chance, we used a more stringent approach of identifying only relationships that were positive and statistically significant at the 5 percent level.
- **Adjusting statistical tests for having examined multiple relationships:** Examining many relationships between practices and growth, as we did in this study, increases the risk of finding some relationships to be statistically significant just by chance. In an alternative approach, we identified relationships that remained positive and statistically significant at the 5 percent level even after accounting for the number of relationships examined by the study.

Among the potentially promising practices identified in Tables III.1 and III.2, we found that certain practices remained potentially promising in many of the alternative approaches, whereas others did not (see Appendix D, Tables D.15 and D.16). In prekindergarten and kindergarten, we identified three tiers of potentially promising practices according to how many alternative approaches continue to identify them as potentially promising:

- **Highest tier:** One practice remains potentially promising in all relevant alternative approaches: focusing on the meaning of texts during pre-reading.
- **Middle tier:** Two practices remain potentially promising in at least half, but not all, alternative approaches: helping students make connections between their prior knowledge and texts and engaging students in defining new words during reading.
- **Lowest tier:** Two practices remain potentially promising in fewer than half of the alternative approaches: focusing on higher-order thinking and focusing on world knowledge.

Likewise, we identified the following three tiers of potentially promising practices in grades 1 to 3:

- **Highest tier:** Two practices remain potentially promising in all relevant alternative approaches: engaging students in defining new words during post-reading and focusing on higher-order thinking.

- **Middle tier:** Two practices remain potentially promising in at least half, but not all, alternative approaches: helping students make connections between their prior knowledge and texts and teaching students to use other comprehension strategies.
- **Lowest tier:** One practice remains potentially promising in fewer than half of the alternative approaches: encouraging students' oral language.

Appendix D provides more details on the methods and results of the alternative approaches.

C. Potentially promising practices for student subgroups

Research indicates that students may respond to certain instructional practices differently based on their familiarity with English (or their home languages) and their comprehension skills. For example, Bowers and Vasilyeva (2011) found that the vocabulary growth of preschool English Learner (EL) students and non-EL students correlated to different aspects of teachers' oral language (quantity and complexity), while Silverman and Crandell (2010) found that the correlation between preschool and kindergarten teachers' practices and students' vocabulary development varied by students' skill levels.

We therefore examined relationships between instructional practices and student growth within subgroups defined by students' home language (English or non-English) and baseline test score (high- or low-achieving).[7] To generate hypotheses on the practices most suited to different types of students, we identified potentially promising practices separately for each subgroup. We did not have the precision to assess whether a practice had statistically different relationships with growth in one subgroup than in another. In fact, no statistical tests were conducted to determine whether the relationships were statistically significantly different across subgroups. Therefore, the finding that a practice is potentially promising in one subgroup but not another does not necessarily mean that its relationship with growth is significantly stronger in the first group than in the second.

The findings, summarized in Tables III.3 and III.4, offer a key lesson for future research: differences in the practices that are potentially promising for students with English and non-English home languages do not closely mirror differences between high- and low-achieving students. For example, in the lower grades, helping students make prior knowledge connections is potentially promising for high achievers but not low achievers, yet it is potentially promising for *both* students with English and non-English home languages. In general, we found little correspondence between patterns based on home language and those based on incoming achievement. A key reason is that many low achievers (38 to 56 percent, depending on the test and grade span) had English as their home language, and many students with non-English home languages (39 to 54 percent) were not low achievers.

[7] High and low achievers were defined as those whose fall test scores on the same measure as the outcome were, respectively, in the top and bottom 40 percent of students in the study who had scores on the measure. We excluded middle-achieving students (those in the middle 20 percent) to consider a sharper contrast in baseline test scores between the high and low achievers. Because we used the fall test score on the same measure as the outcome, the students identified as high achievers for one outcome were not generally the same as those identified as high achievers for a different outcome.

Therefore, in what follows, we discuss the patterns of results separately by home language and incoming achievement (see Appendix D, Tables D.17 through D.31, for detailed results, including results by gender).

Potentially promising practices for subgroups defined by students' home language

Tables III.3 and III.4 present information about which practices were identified as potentially promising for students whose home language was English and for those whose home language was not English. We found that the numbers and types of potentially promising practices differed between the two groups.

- **In the lower grades, more practices were identified as potentially promising for students with a non-English home language compared to those with English as their home language.** In prekindergarten and kindergarten, seven practices were identified as potentially promising for students with a non-English home language; three were identified for students with English as their home language (Table III.3).
- **In the lower grades, different practices were identified as potentially promising for students with a non-English home language compared to those with English as their home language.** World knowledge instruction, two vocabulary-related practices (engaging students in defining new words during pre-reading and reading), and two meaning-related practices (focusing on the meaning of texts during reading and post-reading) were identified as potentially promising for students with non-English home languages, but not for students with English as their home language (Table III.3). On the other hand, focusing on higher-order thinking was identified as potentially promising for students with English as their home language, but not for students with non-English home languages. Only two practices were identified as potentially promising for both subgroups: focusing on the meaning of texts during pre-reading and helping students make connections between their prior knowledge and texts.
- **In the upper grades, similar numbers of practices were identified as potentially promising for students in these two groups.** In grades 1 through 3, six practices were identified as potentially promising for students with English as their home language, and five for students with non-English home languages (Table III.4).
- **In the upper grades, different practices were identified as potentially promising for students in these two groups.** The growth of both groups was positively related to practices focused on the meaning of texts, but the timing of those practices differed (during reading and post-reading for students with non-English home languages, but only during post-reading for students with English as their home language; Table III.4). Similarly, the growth of both groups was positively related to practices focused on defining new words, but the timing of those practices differed (during post-reading for students with non-English home languages, but during and outside of reading for students with English as their home language). Focusing on phonics and grammar and making prior knowledge connections were potentially promising practices for students with non-English home languages, while encouraging students' oral language, teaching students to use other comprehension strategies, and focusing on higher-order thinking were potentially promising practices for students with English as their home language.

Potentially promising practices for subgroups defined by students' skill levels (baseline test scores)

Tables III.3 and III.4 present information about which practices were identified as potentially promising for high- and low-achieving students. We found that the numbers and types of potentially promising practices differed between the two groups.

- **In the lower grades, most of the practices identified as potentially promising for the full sample were also identified as potentially promising for high-achieving students, but just one practice was identified for low-achieving students.** Of the five practices identified as potentially promising in the full prekindergarten and kindergarten sample, all but one—engaging students in defining new words during reading—were also identified for high-achieving students (Table III.3). Only one practice was identified as potentially promising for low-achieving students: focusing on the meaning of texts during pre-reading.
- **In the upper grades, no practices were identified as potentially promising for high-achieving students, but seven were identified for low-achieving students.** Three of these practices were also identified as potentially promising for the full sample: encouraging students' oral language, engaging students in defining new words during post-reading, and teaching students to use other comprehension strategies (Table III.4). Four practices were identified as potentially promising for low-achieving students, but not for the full sample: engaging students in defining new words during reading and outside of reading, and focusing on the meaning of texts during pre-reading and reading.

Table III.3. Instructional practices identified as potentially promising in prekindergarten and kindergarten for student subgroups

Instructional practice	Promising practice?				
	Full sample	English home language	Non-English home language	High achievers[b]	Low achievers[b]
1. Encouraging students' oral language	No	No	No	No	No
2. Focusing on phonics and grammar during reading	No	No	No	No	No
3. Engaging students in defining new words during pre-reading	No	No	Yes	No	No
4. Engaging students in defining new words during reading	Yes	No	Yes	No	No
5. Engaging students in defining new words during post-reading	NA[a]	NA[a]	NA[a]	NA[a]	NA[a]
6. Engaging students in defining new words outside of reading	No	No	No	No	No
7. Focusing on the meaning of texts during pre-reading	Yes	Yes	Yes	Yes	Yes
8. Focusing on the meaning of texts during reading	No	No	Yes	No	No
9. Focusing on the meaning of texts during post-reading	No	No	Yes	No	No
10. Helping students make connections between their prior knowledge and texts	Yes	Yes	Yes	Yes	No
11. Teaching students to use other comprehension strategies	No	No	No	No	No
12. Focusing on world knowledge	Yes	No	Yes	Yes	No
13. Focusing on higher-order thinking	Yes	Yes	No	Yes	No
Number of classrooms	**378**	**346–348**	**203–214**	**292–325**	**286–321**

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The table includes data from 378 prekindergarten and kindergarten classrooms.

A practice is considered potentially promising if there was at least one positive and significant relationship and no negative and significant relationships.

[a] Within the lower grades, we did not find evidence that teachers in the study differed in the usual extent to which they engaged students in defining new words during post-reading. Therefore, we did not examine the relationship between this practice and student growth in the lower grades.

[b] High and low achievers are those whose fall test scores were, respectively, in the top and bottom 40 percent of students in the study.

NA = not applicable.

Table III.4. Instructional practices identified as potentially promising in grades 1 to 3 for student subgroups

Instructional practice	Promising practice?				
	Full sample	English home language	Non-English home language	High achievers[b]	Low achievers[b]
1. Encouraging students' oral language	Yes	Yes	No	No	Yes
2. Focusing on phonics and grammar during reading	No	No	Yes	No	No
3. Engaging students in defining new words during pre-reading	No	No	No	No	No
4. Engaging students in defining new words during reading	No	Yes	No	No	Yes
5. Engaging students in defining new words during post-reading	Yes	No	Yes	No	Yes
6. Engaging students in defining new words outside of reading	No	Yes	No	No	Yes
7. Focusing on the meaning of texts during pre-reading	No	No	No	No	Yes
8. Focusing on the meaning of texts during reading	No	No	Yes	No	Yes
9. Focusing on the meaning of texts during post-reading	No	Yes	Yes	No	No
10. Helping students make connections between their prior knowledge and texts	Yes	No	Yes	No	No
11. Teaching students to use other comprehension strategies	Yes	Yes	No	No	Yes
12. Focusing on world knowledge	No	No	No	No	No
13. Focusing on higher-order thinking	Yes	Yes	No	No	No
Number of classrooms	**220–657**	**199–607**	**112–348**	**163–582**	**184–532**

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

A practice is considered potentially promising if there was at least one positive and significant relationship and no negative and significant relationships.

NA = not applicable.

This page has been left blank for double-sided copying.

IV. DISCUSSION

In this chapter, for each of the practices that this study identified as holding potential promise for improving students' language skills and comprehension in prekindergarten through grade 3, we summarize our findings and briefly discuss some related research. Our purpose is not to provide an exhaustive summary of that previous work—as there are many such reviews available—but to show how our study extends or complements existing research. We then explore whether our findings, in combination with this past evidence, suggest that these practices merit further study. We also identify the remaining gaps in the research that future studies might address. We discuss each of the practices identified as potentially promising for the full sample of students in the study, including encouraging students' oral language (Section A), engaging students in defining new words during reading and post-reading (Section B), focusing on the meaning of texts during pre-reading (Section C), making prior knowledge connections and teaching other comprehension strategies (Section D), focusing on world knowledge (Section E), and focusing on higher-order thinking (Section F). We then summarize and prioritize the suggestions for future research identified throughout this chapter (Section G).

A. Encouraging students' oral language

Dozens of studies have found a positive correlation between oral language development and both listening and reading comprehension (Kendeou et al. 2009; National Early Literacy Panel 2008). This research has fueled interest in strategies for promoting early language growth. Studies have found that it is possible to enhance early language growth by motivating students to use language and then helping them to expand their utterances (Yoder et al. 1995); by reading to students, asking them open-ended questions, and responding in supportive ways to their attempts to answer such questions (Whitehurst et al. 1988); by engaging students in extended discussions during read-alouds (Zucker et al. 2012); and by using syntactically rich talk in the classroom (Huttenlocher et al. 2002). Research also indicates a positive correlation between the degree to which children actively talk with others—particularly with those who are able to extend and clarify the ideas expressed in language—and the sophistication of the language they use (Baker et al. 2006; Gersten et al. 2005).

However, with few exceptions (for example, Dickinson and Porche [2011]), these studies have not yet demonstrated that instructional strategies for promoting early language growth have a direct impact on early reading comprehension. Consequently, this study considered whether teachers' efforts to encourage students' oral language use were related to a range of language and comprehension outcomes. We found an association between encouraging students' oral language and students' background knowledge growth in grade 1. This finding suggests that future research could assess rigorously the effects of techniques to encourage students' language use on a broader range of language and comprehension outcomes.[8]

[8] In the majority of classrooms we observed, teachers spoke with students frequently, used clear and correct language, and often used multiple techniques to encourage students' language. The lack of variation in the frequency and quality of teachers' language use likely restricted our ability to detect further relationships between teachers' language use and other student outcomes.

B. Engaging students in defining new words during reading and post-reading

Research has shown that early vocabulary development can predict later reading comprehension (Hemphill and Tivanan 2008; National Early Literacy Panel 2008; NICHD Early Child Care Research Network 2005; Duff et al. 2015; Song et al. 2015). In addition, research has found that explicit teaching of vocabulary correlates positively to improved reading comprehension (Blachowicz and Fisher 2007; NICHD 2000), as does instruction that increases the amount of exposure children have to the meaning of words (Pressley 2000) and that involves defining words using explicit approaches (Neuman et al. 2011).

However, most experimental studies of the impacts of vocabulary instruction on reading comprehension have focused on students in grades 3 to 8 (NICHD 2000). Studies of instruction in early-grade classrooms have suggested that most growth in young children's vocabulary may be due to informal means (including media and interactions with parents) rather than to formal instruction (Biemiller 2003). Nevertheless, there are many studies of the effects of vocabulary instruction on vocabulary improvement in preschool and the primary grades (Beck and McKeown 2007; Collins 2010; Elley 1989; Penno et al. 2002; Sénéchal 1997). None of those early-grade studies have determined whether the effects carry over to listening comprehension, reading comprehension, or background knowledge, or even to improvements in other aspects of oral language.

In this study, we found a mix of positive and negative relationships between vocabulary instruction and student outcomes, including at least some positive relationships with basic language skills, background knowledge, and reading comprehension. This finding suggests that further rigorous study is needed to determine the effects of vocabulary instruction in the early grades on a range of language and comprehension outcomes, beyond just vocabulary development.

C. Focusing on the meaning of texts during pre-reading

Reading comprehension requires attention to the meaning of a text. Even adults who are good readers experience mind-wandering during reading (McVay and Kane 2012), and attending to meaning can be an even greater challenge for young children (Cain and Bignell 2014). Young children can become so focused on decoding and fluency during reading that they lose the meaning of the text, undermining comprehension. Instructional practices can either guide students' attention toward meaning or distract from it (Anderson et al. 1991).

Research has found that activities that focus students' attention on meaning before, during, and after reading are positively correlated with comprehension. For example, studies have shown that focusing on meaning before reading—such as by introducing the topic, asking questions, identifying the purpose of the text, or encouraging predictions—is positively associated with students' comprehension (Lewis and Mensink 2012; Mills et al. 1995; Spires et al. 1992). Similarly, research has found that asking meaning-focused questions during and after reading is positively correlated with students' comprehension (Casteel 1993; Koskinen et al. 1989; Law 2008; NICHD 2000; Shannon et al. 1988; van den Broek 1990; Williams et al. 2005). However, most of these studies examined the comprehension of older students (NICHD 2000; Shanahan et al. 2010).

This study helps fill those gaps by examining whether focusing on meaning is related to the overall language and comprehension growth of younger children. As reported above, we found mixed results. A focus on grammar or phonics during text reading lessons was not positively associated with any of the outcomes of interest, suggesting that those kinds of considerations during reading may be distracting from meaning. At the same time, the analyses showed that heightened emphasis on meaning during these lessons had somewhat inconsistent relationships with language and comprehension outcomes, with a mix of positive, negative, and null relationships depending on the grade span, outcome, and timing of the emphasis on meaning (before, during, or after the text was read). Only one of the meaning-focused practices—focusing on the meaning of texts before reading—had a positive relationship with an outcome (basic language skills of prekindergarten and kindergarten students) and no negative relationships in the same grade span.

Given the mixed findings obtained here, and the clear gaps in the research, further research on the practice of focusing on meaning is needed. Further studies could rigorously examine the effects of meaning-focused reading instruction on overall language and comprehension skills in the early grades.

D. Making prior knowledge connections and teaching other comprehension strategies

Research on prior knowledge connections indicates that teaching students to think about what they already know about a topic can improve reading comprehension (Hansen 1981) and that this strategy can be combined effectively with the use of other strategies in supporting students' comprehension (Shanahan et al. 2010). In addition, research indicates that the use of other reading comprehension strategies is associated with gains in students' reading and listening comprehension (NICHD 2000), even with younger students (Shanahan et al. 2010). Studies have demonstrated the benefits of the following comprehension strategies, either individually or in combination: activating prior knowledge, predicting, purpose-setting, questioning, visualizing, self-monitoring, summarizing, story mapping, identifying text structures, interpreting cohesion clues, reflecting on the author's purpose, and thinking aloud (Beck et al. 1997; Brown et al. 1996; Duffy et al. 1986; Eilers and Pinkley 2006; Kinnunen and Vauras 1995; Spörer et al. 2009; Williams et al. 2005).

Despite this body of research, which shows effects of reading strategy instruction in small numbers of classrooms, there is little information available about the effectiveness of such instruction when used on a large scale. This study has contributed large-scale exploratory evidence showing that helping students make prior knowledge connections was positively associated with listening comprehension growth in prekindergarten and kindergarten and with growth in reading comprehension and background knowledge in grades 1 to 3. In addition, teaching students to use other comprehension strategies, such as predicting, summarizing, and questioning, was positively associated with listening comprehension growth in grades 1 to 3.

Given these findings and those of past research, a next step would be for research to explore the effect of promoting and maintaining instruction in reading comprehension strategies on a large scale. In the current policy context, the emphasis on "close reading" in the Common Core State Standards (National Governors Association Center for Best Practices 2010) has

promoted instructional practices that focus student attention on the information in a text, without regard for other knowledge students may bring to the text. Educators need more evidence on the relative value of close reading versus using prior knowledge more intentionally in promoting comprehension growth in the early grades.

E. Focusing on world knowledge

Extending the notion that students' prior knowledge can support comprehension, some theorists argue that children would benefit from concerted efforts to develop their "knowledge of the world" (Hirsch 1996, 2003, 2006). World knowledge refers to awareness of general aspects of daily life (for example, how time is measured and what jobs people do) and specific cultural, historical, and scientific information (for example, George Washington was the first U.S. president, gravity makes the apple fall, and Little Red Riding Hood meets a wolf). Synonyms for world knowledge include domain knowledge, declarative knowledge, core knowledge, crystallized intelligence, funds of information, and cultural literacy.

Research has found positive relationships between world knowledge and reading comprehension (Benson 2008; Flanagan 2000; Nation et al. 2002), but it is unclear whether such knowledge supports reading comprehension or whether reading is just a particularly effective avenue to increased world knowledge. Nevertheless, research has supported the idea that both the amount and type of subject-matter knowledge are associated with growth in comprehension, at least with older students (Alexander et al. 1994; Kozminsky and Kozminsky 2001; Reynolds and Turek 2012). A prominent example of a world knowledge curriculum is the one developed by the Core Knowledge Foundation (1999), and studies of its impact on reading comprehension have found mixed results (Sterbinsky et al. 2006; Stringfield et al. 2000).

The question of whether world knowledge instruction contributes to the development of students' reading comprehension skills is particularly timely because efforts to increase test performance in reading have led many schools to narrow the curriculum, teaching less core content in social studies and science (David 2011). This study found that information-rich instruction was related to more growth in basic language skills in prekindergarten and kindergarten, but less growth in basic language skills in grades 1 to 3; and at neither level was world knowledge related to listening or reading comprehension growth.

The extent to which world knowledge instruction promotes language and comprehension growth may depend on whether the content of the instruction goes beyond what students already know. Future research should examine deliberate strategies for teaching world knowledge that go beyond students' prior knowledge and assess the impacts of these strategies on language and comprehension outcomes.

F. Focusing on higher-order thinking

A common instructional practice for increasing the intellectual rigor of reading is to ask questions that require higher-order thinking or questions that ask students to explain their reasoning about a text. Research has found positive correlations between students' reading comprehension and teachers' use of higher-order questions (Andre 1979; Franks 1996; Taylor et al. 2000). Similarly, studies have found that students who are better at answering inferencing questions (Lepola et al. 2012; Tompkins et al. 2013) or who engage in metacognitive acts, such

as explaining their reasoning, demonstrate better reading comprehension (Erlich et al. 1999; Segers and Verhoeven 2016).

Although most studies of higher-order thinking focus on older students (upper elementary and secondary), some studies have found positive correlations between encouraging students' higher-order thinking and reading comprehension in the early elementary grades (Cain and Oakhill 1999; McGee and Johnson 2003). This study adds to the evidence base for prekindergarten and the early elementary grades, finding that instructional practices that fostered higher-order thinking had positive relationships with basic language skills growth in prekindergarten and kindergarten and with background knowledge growth in grade 1. Given this finding, future research should examine instructional practices aimed at increasing the intellectual rigor of student work in the early grades to measure the effects of those practices on language and comprehension outcomes.

G. Summary and prioritization of suggestions for future research

The practices this study has identified as potentially promising deserve further rigorous research. For each of the potentially promising practices, this chapter has highlighted important gaps in our knowledge about whether and how these practices promote students' language development and comprehension.

Given that this study has offered several suggestions for future research, researchers may need to determine an order of priority for topics to be examined. For example, the research community could give higher priority to examining practices that are more consistently identified as potentially promising across multiple analytic approaches. In Chapter III, Section B, we classified the potentially promising practices in each grade span into three tiers—high, middle, or low—according to the consistency with which they were identified as potentially promising by alternative analytic approaches. To the extent that future research studies include both grade spans together, the highest-priority practices would be those that this study has consistently identified as potentially promising across multiple approaches in both grade spans.

Based on these principles for prioritizing research areas, further research on vocabulary instruction would merit high priority, given that at least one vocabulary-oriented practice was in either the highest or middle tier in each grade span. Research on prior knowledge connections (middle tier in both grade spans), promoting higher-order thinking (highest tier in the upper grades and lowest tier in the lower grades), and focusing on the meaning of texts (highest tier in the lower grades) would also receive relatively high priority. The full list of topics for future research identified in this chapter, in a suggested order of priority, is as follows:

- The effects of vocabulary instruction on a range of language and comprehension outcomes, beyond just vocabulary development
- The relative value of close reading (focusing on the information in a text) versus using prior knowledge more intentionally to promote reading comprehension growth
- The effects of instruction that promotes higher-order thinking on students' language development and comprehension

- The extent to which focusing on the meaning of texts in reading lessons promotes overall language and comprehension skills, not just comprehension of the specific texts being taught
- The effects of promoting and maintaining instruction in reading comprehension strategies on a large scale
- The effects of encouraging student's oral language on a range of language and comprehension outcomes, beyond just oral language skills
- The extent to which deliberate strategies to teach world knowledge (that go beyond students' prior knowledge) lead to growth in language and comprehension

This study's findings also suggest that future research on these potentially promising practices should carefully examine the ways in which these practices have different effects on students with different home language backgrounds or skill levels. As shown in Chapter III, Section C, most practices identified as potentially promising for the full student sample in each grade span were positively related to the growth of either students who spoke English at home or students who did not—but typically not both. Likewise, most potentially promising practices in this study were positively related to the growth of either high achievers or low achievers, but usually not both. These results suggest that when investigating the research topics listed above, researchers should determine the extent to which results differ across students of different language backgrounds and/or skill levels. Answering these questions will require obtaining sufficiently large samples of students to ensure adequate statistical power for examining the effects of these practices on student subgroups.

Finally, given that this study's classroom observations could not capture some aspects of instruction that might influence student growth, researchers may consider examining those aspects further. As discussed in Chapter II, Section E, this study was not able to examine the extent to which student growth was related to (1) the degree to which teachers connected and integrated lessons across the school day or days, (2) the frequency with which teachers changed classroom structures and the ways in which teachers' practices were influenced by such structures, (3) the content and difficulty level of the subject matter taught, (4) the quality and difficulty level of the instructional materials, (5) the degree to which instruction was planned rather than spontaneous, (6) the quality of teachers' language, and (7) the quantity and quality of peers' language. Moreover, given that this study took place in large districts with many high- and low-performing Title I schools, there is also a need to replicate the study in smaller, non-urban districts. These topics may be appropriate for future research.

REFERENCES

Alexander, P.A., J. M. Kulikowich, and S. K. Schulze. "How Subject-Matter Knowledge Affects Recall and Interest." *American Educational Research Journal*, vol. 31, 1994, pp. 313–337.

Anderson, R. C., I. A. G. Wilkinson, and J. M. Mason. "A Microanalysis of the Small-Group, Guided Reading Lesson: Effects of an Emphasis on Global Story Meaning." *Reading Research Quarterly*, vol. 26, 1991, pp. 417–441.

André, T. "Does Answering Higher Level Questions While Reading Facilitate Productive Learning?" *Review of Educational Research*, vol. 49, 1979, pp. 280–319.

Aud, Susan, Sidney Wilkinson-Flicker, Paul Kristapovich, Amy Rathbun, Xiaolei Wang, and Jijun Zhang. *The Condition of Education 2013* (NCES 2013-037). Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2013.

Baker, S., R. Gersten, D. Haager, and M. Dingle. "Teaching Practice and the Reading Growth of First-Grade English Learners: Validation of an Observation Instrument." *Elementary School Journal*, vol. 107, 2006, pp. 199–219.

Bandeira de Mello, Victor, Charles Blankenship, and Don McLaughlin. "Mapping State Proficiency Standards Onto NAEP Scales: 2005–2007." Research and Development Report, NCES 2010-456. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2009.

Bassok, Daphna, Scott Latham, and Anna Rorem. "Is Kindergarten the New First Grade?" *AERA Open* vol. 1, no. 4, 2016, pp. 1–31.

Beck, I.L., and M. McKeown. "Conditions of Vocabulary Acquisition." In *Handbook of Reading Research, vol. 2,* edited by R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson. White Plains, NY: Longman, 1991, pp. 789–814.

Beck, I., and M. McKeown. "Increasing Young Low-Income Children's Oral Vocabulary Repertories through Rich and Focused Instruction." *Elementary School Journal*, vol. 107, 2007, pp. 251–271.

Beck, I. L., M. G. McKeown, R. L. Hamilton, and L. Kugan. *Questioning the Author: An Approach for Enhancing Student Engagement with Text*. Newark, DE: International Reading Association, 1997.

Benjamini, Y., and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society Series B Methodological*, vol. 57, no. 1, 1995, pp. 289–300.

Benson, N. "Cattell-Horn-Carroll Cognitive Abilities and Reading Achievement." *Journal of Psychoeducational Assessment*, vol. 26, 2008, pp. 27–41.

Berko Gleason, J. (ed.). *The Development of Language.* 6th ed. Boston: Allyn and Bacon, 2005.

Biemiller, A. "Vocabulary: Needed if More Children Are to Read Well." *Reading Psychology*, vol. 24, 2003, pp. 323–335.

Blachowicz, C. L. Z., and P. Fisher. "Best Practices in Teaching Vocabulary." In L. Gambrell, L. Morrow, and M. Pressley, *Best Practices in Literacy Instruction*. 3rd ed. New York: Guilford, 2007, pp. 178–203.

Bowers, E. P., and M. Vasilyeva. "The Relation between Teacher Input and Lexical Growth of Preschoolers." *Applied Psycholinguistics,* vol. 32, no. 1, 2011, pp. 221–241.

Brassard, Marla R., and Ann E. Boehm. *Preschool Assessment: Principles and Practices*. New York: Guilford Press, 2007.

Brown, R., M. Pressley, P. Van Meter, and T. Schuder. "A Quasi-Experimental Validation of Transactional Strategies Instruction with Previously Low-Achieving Second-Grade Readers." *Journal of Educational Psychology*, vol. 88, 1996, pp. 18–37.

Cain, K., and S. Bignell. "Reading and Listening Comprehension and Their Relation to Inattention and Hyperactivity." *British Journal of Educational Psychology*, vol. 84, 2014, pp. 108–124.

Cain, K., and J. V. Oakhill. "Inference Making Ability and Its Relation to Comprehension Failure in Young Children." *Reading and Writing*, vol. 11, 1999, pp. 489–503.

Carlisle, Joanne F., and Melinda S. Rice. *Improving Reading Comprehension: Research-Based Principles and Practices*. Baltimore: York Press, Inc., 2002.

Casteel, M. "Effects of Inferences Necessity and Reading Goal on Children's Inferential Integration." *Journal of Educational Psychology*, vol. 88, 1993, pp. 484–507.

Cattell, R. B. "The Scree Test for the Number of Factors." *Multivariate Behavioral Research*, vol. 1, 1966, pp. 245–276.

Catts, Hugh W., M. E. Fey, J. Bruce Tomblin, and X. Zhang. "A Longitudinal Investigation of Reading Outcomes in Children with Language Impairments." *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 6, 2002, pp. 1142–1157.

Chan, K.S., and P.G. Cole. "An Aptitude-Treatment Interaction in a Mastery Learning Model of Instruction." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April 1986.

Child Trends Databank. "Full-day Kindergarten." February 2015. Available at http://www.childtrends.org/wp-content/uploads/2013/06/102_Full-day-kindergarten.pdf.

Claessens, Amy, Greg J. Duncan, and Mimi Engel. "Kindergarten Skills and Fifth-Grade Achievement: Evidence from the ECLS-K." *Economics of Education Review,* vol. 28, no. 4, 2009, pp. 415–427.

Collins, M. F. "ELL Preschoolers' English Vocabulary Acquisition from Storybook Reading." *Early Childhood Research Quarterly*, vol. 25, 2010, pp. 84–97.

Core Knowledge Foundation. *Core Knowledge Sequence: Content Guidelines for Grades K-8.* Charlottesville, VA: Author, 1999.

Correnti, R., and B. Rowan. "Opening Up the Black Box: Literacy Instruction in Schools Participating in Three Comprehensive School Reform Programs." *American Educational Research Journal,* vol. 44, 2007, pp. 298–338.

Crissey, Sarah R. "Educational Attainment in the United States: 2007." Current Population Reports, P20-560. Washington, DC: U.S. Department of Commerce, January 2009.

David, J. L. "Research Says… / High-Stakes Testing Narrows The Curriculum." *Educational Leadership*, vol. 68, no. 6, 2011, pp. 78–80.

Dickinson, D. K., and M. V. Porche. "Relation Between Language Experiences in Preschool Classrooms and Children's Kindergarten and Fourth-Grade Language and Reading Abilities." *Child Development*, vol. 82, no. 3, 2011, pp. 870–886.

Duff, F. J., G. Reen, K. Plunkett, and K. Nation. "Do Infant Vocabulary Skills Predict School-Age Language and Literacy Outcomes?" *Journal of Child Psychology and Psychiatry*, vol. 56, 2015, pp. 848–856.

Duffy, G., L. Roehler, M. Meloth, L. Vavrus, C. Book, J. Putnam, and R. Wesselman. "The Relationship Between Explicit Verbal Instruction During Reading Skill Instruction and Student Awareness and Achievement: A Study Of Reading Teacher Effects." *Reading Research Quarterly*, vol. 21, 1986, pp. 237–252.

Duncan, Greg, Chantelle J. Dowsett, Amy Claessens, Katherine Magnuson, Aletha C. Huston, Pamela Klebanov, Linda Pagani, Leon Feinstein, Mimi Engel, Jeanne Brooks-Gunn, Holly Sexton, Kathryn Duckworth, and Crista Japel. "School Readiness and Later Achievement." *Developmental Psychology*, vol. 43, no. 6, 2007, pp. 1428–1446.

Duncan, Sharon E., and Edward DeAvila. *Preschool Language Assessment Survey 2000 Examiner's Manual: English Forms C and D.* Monterey, CA: CTB/McGraw-Hill, 1998.

Durso, F.T., and K.A. Coggins. "Organized Instruction for the Improvement of Word Knowledge Skills." *Journal of Educational Psychology,* vol. 83, 1991, pp. 109–111.

Eilers, L. H., and C. Pinkley. "Metacognitive Strategies Help Students to Comprehend All Text." *Reading Improvement*, vol. 43, 2006, pp. 13–29.

Elley, W. "Vocabulary Acquisition from Listening to Stories." *Reading Research Quarterly*, vol. 24, 1989, pp. 174–187.

Erlich, M. F., M. Remond, and H. Tardieu. "Processing of Anaphoric Devices in Young Skilled and Less Skilled Comprehenders: Differences in Metacognitive Monitoring." *Reading and Writing*, vol. 11, 1999, pp. 29–63.

Fabrigar, Leandre R., Duane T. Wegener, Robert C. MacCallum, and Erin J. Strahan. "Evaluating the Use of Exploratory Factor Analysis in Psychological Research." *Psychological Methods*, vol. 4, no. 3, 1999, pp. 272–299.

Flanagan, D. P. "Wechsler-Based CHC Cross-Battery Assessment and Reading Achievement: Strengthening the Validity of Interpretations Drawn From Wechsler Test Scores." *School Psychology Quarterly*, vol. 15, 2000, pp. 295–329.

Francis, D. J., J. M. Fletcher, H. W. Catts, and J. B. Tomblin. "Dimensions Affecting the Assessment of Reading Comprehension." In *Children's Reading Comprehension and Assessment,* edited by S. G. Paris and S. A. Stahl. Mahwah, NJ: Erlbaum, 2005.

Franks, B. A. "Deductive Reasoning in Narrative Contexts: Developmental Trends and Reading Skill Effects." *Genetic, Social, and General Psychology Monographs*, vol. 122, 1996, pp. 75–105.

Gagne, E.D., and D. Memory. "Instructional Events and Comprehension: Generalization Across Passages." *Journal of Reading Behavior,* vol. 10, no. 4, 1978, pp. 321–335.

Gamse, Beth C., Howard S. Bloom, James J. Kemple, and Robin Tepper Jacob. "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, 2008a. Available at http://ies.ed.gov/ncee/pubs/20084016.

Gamse, Beth C., Robin Tepper Jacob, Megan Horst, Beth Boulay, and Fatih Unlu. "Reading First Impact Study Final Report" (NCEE 2009-4038). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2008b. Available at http://ies.ed.gov/ncee/pubs/20094038.

Gersten, R., S. K. Baker, D. Haager, and A. Graves. "Exploring the Role of Teacher Quality in Predicting Reading Outcomes for First-Grade English Learners: An Observational Study." *Remedial and Special Education*, vol. 26, 2005, 197–206.

Glenn-Applegate, K., A. Breit-Smith, L.M. Justice, and S.B. Piasta. "Artfulness in Young Children's Spoken Narratives." *Early Education and Development*, vol. 21, no. 3, 2010, pp. 468–493.

Goldman, S.R., and C.K. Varnhagen. "Memory for Embedded and Sequential Story Structures." *Journal of Memory and Language,* vol. 25, 1986, pp. 401–418.

Guarino, Cassandra M., Mark D. Reckase, and Jeffrey M. Wooldridge "Can Value-Added Measures of Teacher Performance Be Trusted?" *Education Finance and Policy*, vol. 10, no. 1, 2014, pp. 117–156.

Hansen, J. "The Effects of Inference Training and Practice on Young Children's Reading Comprehension." *Reading Research Quarterly*, vol. 16, no. 3, 1981, pp. 391–417.

Hemphill, L., and T. Tivanan. "The Importance of Early Vocabulary for Literacy Achievement in High-Poverty Schools." *Journal of Education for Students Placed at Risk*, vol. 13, 2008, pp. 426–451.

Hendrickson, A. E., and P. O. White. "Promax: A Quick Method for Rotation to Oblique Simple Structure." *British Journal of Statistical Psychology*, vol. 17, pp. 65–70.

Hirsch, E. D. *The Schools We Need and Why We Don't Have Them.* New York, NY: Doubleday Dell Publishing Group, Inc., 1996.

Hirsch, E. D. "Reading Comprehension Requires Knowledge—of Words and the World." *American Educator*, vol. 27, no. 1, 2003, pp. 10–13, 16–22, 28–29, 48.

Hirsch, E. D. *The Knowledge Deficit: Closing the Shocking Education Gap for American Children.* New York: Houghton Mifflin, 2006.

Hoffman, James, Misty Sailors, Gerald. R. Duffy, and S. Natasha Beretvas. "The Effective Elementary Classroom Literacy Environment: Examining the Validity of the TEX-IN3 Observation System." *Journal of Literacy Research,* vol. 36, no. 3, 2004, pp. 303–334.

Hoover, W. A., and P. B. Gough. "The Simple View of Reading." *Reading and Writing: An Interdisciplinary Journal,* vol. 2, 1990, pp. 127–160.

Houston, W. M., M. R. Raymond, and J. C. Svec. "Adjustments for Rater Effects in Performance Assessment." *Applied Psychological Measurement*, vol. 15, no. 4, 1991, pp. 409–421.

Huber, P. J. "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1967, pp. 221–233.

Huttenlocher, J., M. Vasilyeva, E. Cymerman, and S. Levine. "Language Input and Child Syntax." *Cognitive Psychology*, 45, 2002, 337–374.

Jackson, Russell, Ann McCoy, Carol Pistorino, Anna Wilkinson, John Burghardt, Melissa Clark, Christine Ross, Peter Schochet, and Paul Swank. "National Evaluation of Early Reading First: Final Report." U.S. Department of Education, Institute of Education Sciences. Washington, DC: U.S. Government Printing Office, 2007.

Jarmulowicz, L., S. E. Hay, V. L. Taran, and C. A. Ethington. "Fitting Derivational Morphophonology into a Developmental Model of Reading." *Reading and Writing*, vol. 21, 2008, pp. 275–297.

Judkins, David, Robert St. Pierre, Babette Gutmann, Barbara Goodson, Adrienne von Glatz, Jennifer Hamilton, Ann Webber, Patricia Troppe, and Tracy Rimdzius. "A Study of Classroom Literacy Interventions and Outcomes in Even Start." U.S. Department of Education, Institute for Education Sciences, NCEE 2008-4029, 2008.

Justice, L. M., A. S. McGinty, S. Q. Cabell, C. R. Kilday, K. Knighton, and G. Huffman. "Language and Literacy Curriculum Supplement for Preschoolers Who Are Academically At Risk: A Feasibility Study." *Language, Speech, and Hearing Services in Schools*, vol. 41, 2010, pp. 161–178.

Keenan, Janice M., Rebecca S. Betjemann, and Richard K. Olson. "Reading Comprehension Tests Vary in the Skills They Assess: Differential Dependence on Decoding and Oral Comprehension." *Scientific Studies of Reading*, vol. 12, no. 3, 2008, pp. 281–300.

Kelly, M., D.W. Moore, and B.F. Tuck. "Reciprocal Teaching in a Regular Primary School Classroom." *Journal of Educational Research,* vol. 88, 1994, pp. 53–61.

Kendeou, P., P. van den Broek, M. J. White, and J. S. Lynch. "Predicting Reading Comprehension in Early Elementary School: The Independent Contributions of Oral Language and Decoding Skills." *Journal of Educational Psychology*, vol. 101, 2009, pp. 765–778.

Kinnunen, R., and M. Vauras. "Comprehension Monitoring and the Level of Comprehension in High- and Low-Achieving Primary School Children's Reading." *Learning and Instruction*, vol. 5, 1995, pp. 153–165.

Koskinen, P. S., L. B. Gambrell, and B. A. Kapinus. "The Effects of Repeated Reading and Retelling on Young Children's Reading Comprehension." In S. McCormick and J. Zutell (Eds.), *Cognitive and Social Perspectives for Literacy Research and Instruction*. Chicago, IL: National Reading Conference, 1989, pp. 233–240.

Kozminsky, E., and L. Kozminsky. "How Do General Knowledge and Reading Strategies Ability Relate to Reading Comprehension of High School Students at Different Educational Levels?" *Journal of Research in Reading*, vol. 24, 2001, pp. 187–204.

Landry, S.H., A. Crawford, S. Gunnewig, and P.R. Swank. *Teacher Behavior Rating Scale.* Unpublished research instrument. Houston, TX: University of Texas Health Science Center at Houston, Center for Improving the Readiness of Children for Learning and Education, 2001.

Law, Y. "Effects of Cooperative Learning on Second Graders' Learning from Text." *Educational Psychology*, vol. 28, 2008, pp. 567–582.

Lepola, J., J. Lynch, E. Laakkonen, M. Silven, and P. Niemi. "The Role of Inference Making and Other Language Skills in the Development of Narrative Listening Comprehension in 4-6-Year-Old Children." *Reading Research Quarterly*, vol. 47, 2012, pp. 259–282.

Lewis, M. R., and M. C. Mensink. "Prereading Questions and Online Text Comprehension." *Discourse Processes*, vol. 49, 2012, pp. 367–390.

Liang, Kung-Yee, and Scott L. Zeger. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, vol. 73, 1986, pp. 13–22.

Lipsey, M. W., D. C. Farran, and K. G. Hofer. *A Randomized Control Trial of the Effects of a Statewide Voluntary Prekindergarten Program on Children's Skills and Behaviors Through Third Grade* (Research Report). Nashville, TN: Vanderbilt University, Peabody Research Institute, 2015.

Luo, W., and O. M. Kwok. "The Impacts of Ignoring a Crossed Factor in Analyzing Cross-Classified Data." *Multivariate Behavioral Research*, vol. 44, no. 2, 2009, pp. 182–212.

Lyster, S. A. H., A. O. Lervåg, and C. Hulme. "Preschool Morphological Training Produces Long-term Improvements in Reading Comprehension." *Reading and Writing,* vol. 29, 2016, pp. 1269–1288.

MacCallum, R. C., S. Zhang, K. J. Preacher, and D. D. Rucker. "On the Practice of Dichotomization of Quantitative Variables." *Psychological Methods*, vol. 7, 2002, 19–40.

Marley, S.C., J.R. Levin, and A.M. Glenberg. "Improving Native American Children's Listening Comprehension Through Concrete Representations." *Contemporary Educational Psychology,* vol. 32, 2007, pp. 537–550.

Marsh, H. W., A. J. Morin, P. D. Parker, and G. Kaur. "Exploratory Structural Equation Modeling: An Integration of the Best Features of Exploratory and Confirmatory Factor Analysis." *Annual Review of Clinical Psychology*, vol. 10, 2014, 85-110.

Mashburn, A., L. M. Justice, A. McGinty, and L. Slocum. "The Impacts of a Scalable Intervention on the Language and Literacy Development of Rural Pre-Kindergarteners." *Applied Developmental Science*, vol. 20, no. 1, 2016, pp. 61–78.

McGee, A., and H. Johnson. "The Effect of Inference Training on Skilled and Less Skilled Comprehenders." *Educational Psychology*, vol. 23, 2003, pp. 49–59.

McGrew, Kevin S., Fredrick A. Schrank, and Richard W. Woodcock. *Woodcock-Johnson III Normative Update Technical Manual.* Rolling Meadows, IL: Riverside Publishing, 2007.

McGrew, Kevin S., and Richard W. Woodcock. *Woodcock-Johnson III Technical Manual.* Itasca, IL: Riverside Publishing, 2001.

McKeown, M.G., I.L. Beck, and R.G.K. Blake. "Rethinking Reading Comprehension Instruction: A Comparison of Instruction for Strategies and Content Approaches." *Reading Research Quarterly,* vol. 44, no. 3, 2009, pp. 218–255.

McVay, J. C., and M. J. Kane. "Why Does Working Memory Capacity Predict Variation in Reading Comprehension? On the Influence of Mind Wandering and Executive Attention." *Journal of Experimental Psychology: General*, vol. 141, 2012, pp. 302–320.

Meyers, J. L. and S. N. Beretvas. "The Impact of Inappropriate Modeling of Cross-Classified Data Structures." *Multivariate Behavioral Research*, vol. 41, no. 4, 2006, pp. 473–497.

Mills, C. B., V. A. Diehl, D. P. Birkmire, and L. C. Mou. "Reading Procedural Texts: Effects of Purpose for Reading and Predictions of Reading Comprehension Models." *Discourse Processes*, vol. 20, 1995, pp. 79–107.

Morris, C. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association*, vol. 78, no. 381, 1983, pp. 47–55.

Murphy, P.K., I.A.G. Wilkinson, A.O. Soter, M.N. Hennessey, and J.F. Alexander. "Examining the Effects of Classroom Discussion on Students' Comprehension of Text: A Meta-Analysis." *Journal of Educational Psychology,* vol. 101, no. 3, 2009, pp. 740–764.

Muthén, B.O. "Multilevel Factor Analysis of Class and Student Achievement Components." *Journal of Educational Measurement*, vol. 28, 1991, pp. 338–354.

Nagy, W.E., R.C. Anderson, and P.A. Herman, "Learning Word Meanings from Context During Normal Reading." *American Educational Research Journal, vol. 24,* 1987, pp. 237–270.

Nation, K., P. Clarke, and M. J. Snowling. "General Cognitive Ability in Children with Reading Comprehension Difficulties." *British Journal of Educational Psychology*, vol. 72, 2002, pp. 549–560.

Nation, Kate, and Margarte J. Snowling. "Assessing Reading Difficulties: The Validity and Utility of Current Measures of Reading Skill." *British Journal of Educational Psychology*, vol. 67, 1997, pp. 359–370.

National Early Literacy Panel. *Developing Early Literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy, 2008.

National Governors Association Center for Best Practices, Council of Chief State School Officers. *Common Core State Standards*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010.

National Institute of Child Health and Human Development (NICHD). *Report of the National Reading Panel. Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction*. Rockville, MD: National Institute of Child Health and Human Development, 2000. Available at http://www.nichd.nih.gov/publications/nrp/smallbook.cfm.

National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. "Pathways to Reading: The Role of Oral Language in the Transition to Reading." *Developmental Psychology*, vol. 41, 2005, pp. 428–442.

Nelson, K. "Structure and Strategy in Learning to Talk." *Monographs of the Society for Research in Child Development,* vol. 38, nos. 1–2, serial no. 149, 1973.

Neuman, S.B. "Enhancing Children's Comprehension Through Previewing." *National Reading Conference Yearbook,* vol. 37, 1988, pp. 219–224.

Neuman, S., E. Newman, and E. Dwyer. "Educational Effects of a Vocabulary Intervention on Preschoolers' Word Knowledge and Conceptual Development." *Reading Research Quarterly,* vol. 46, no. 3, 2001, pp. 249–272.

Nicholson, T., and S. Schachter. "Spelling Skill and Teaching Practice—Putting Them Back Together Again." *Language Arts,* vol. 56, no. 7, 1979, pp. 804–809.

Penno, J. F., A. G. Wilkinson, and D. W. Moore. "Vocabulary Acquisition from Teacher Explanation and Repeated Listening to Stories: Do They Overcome the Matthew Effect?" *Journal of Educational Psychology*, vol. 94, 2002, pp. 23–33.

Pollack, Judith M., Sally Atkins-Burnett, Donald Rock, and Michael J. Weiss. "Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS–K), Psychometric Report for the Third Grade" (NCES 2005–062). Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2005.

Pianta, R., C. Howes, M. Burchinal, D. Bryant, R. M. Clifford, D. M. Early, and O. Barbarin. "Features of Pre-Kindergarten Programs, Classrooms, and Teachers: Prediction of Observed Classroom Quality and Teacher-Child Interactions." *Applied Developmental Science*, vol. 9, no. 3, 2005, pp. 144–159.

Pianta, Robert C., Karen M. LaParo, and Bridget K. Hamre. *Classroom Assessment Scoring System Manual, K-3 Version.* Charlottesville, VA: Center for Advanced Study of Teaching and Learning, 2006.

Preschool Curriculum Evaluation Research (PCER) Consortium. "Effects of Preschool Curriculum Programs on School Readiness." Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Research, 2008.

Pressley, M. "Comprehension Instruction: What Makes Sense Now, What Might Make Sense Soon." In M. L. Kamil, P. Mosenthal, P. D. Pearson, and R. Barr (Eds.). *Handbook of Reading Research: Volume III*. New York: Longman, 2000.

Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Paedagogiske Institute, 1960.

Raudenbush, Stephen W., Andres Martinez, Howard Bloom, Pei Zhu, and Fen Lin. *Studying the Reliability of Group-Level Measures with Implications for Statistical Power: A Six-Step Paradigm.* Unpublished manuscript. Chicago: University of Chicago, 2011.

Raymond, M. R. and C. Viswesvaran. "Least Squares Models to Correct for Rater Effects in Performance Assessment." *Journal of Educational Measurement*, vol. 30, no. 3, 1993, pp. 253–268.

Reynolds, M. R., and J. J. Turek. "A Dynamic Developmental Link Between Verbal Comprehension-Knowledge (Gc) and Reading Comprehension: Verbal Comprehension Knowledge Drives Positive Change in Reading Comprehension." *Journal of School Psychology*, vol. 50, 2012, pp. 841–863.

Rosenshine, B., C. Meister, and S. Chapman. "Teaching Students to Generate Questions: A Review of the Intervention Studies." *Review of Educational Research,* vol. 66, no. 2, 1996, pp. 181–221.

Royston, P., D. G. Altman, and W. Sauerbrei. "Dichotomizing Continuous Predictors in Multiple Regression: A Bad Idea." *Statistics in Medicine*, vol. 25, 2006, pp. 127–141.

Ruddell, R.B. "Developing Comprehension Abilities: Implications from Research for an Instructional Framework." In *What Research Has to Say About Reading Instruction,* edited by S.J. Samuels. Newark, DE: International Reading Association, 1978, pp. 109–120.

Scott, Kathleen A., Jenny A. Roberts, and Rena Krakow. "Oral and Written Language Development of Children Adopted from China." *American Journal of Speech-Language Pathology*, vol. 17, 2008, pp. 150–160.

Segers, E., and L. Verhoeven. "How Logical Reasoning Mediates the Relation Between Lexical Quality and Reading Comprehension." *Reading and Writing*, vol. 29, 2016, pp. 577–590.

Semel, Eleanor, Elisabeth H. Wiig, and Wayne A. Secord. *Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF–4)*. San Antonio, TX: Pearson Assessments, 2003.

Sénéchal, M. "The Differential Effect of Storybook Reading on Preschoolers' Acquisition of Expressive and Receptive Vocabulary." *Journal of Child Language*, vol. 24, 1997, pp. 123–138.

Shanahan, T., K. Callison, C. Carriere, N. K. Duke, P. D. Pearson, C. Schatschneider, and J. Torgesen. *Improving Reading Comprehension in Kindergarten Through 3rd Grade: A Practice Guide* (NCEE 2010-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2010. Available at whatworks.ed.gov/publications/practiceguides.

Shannon, P. W., E. J. Kameenui, and J. F. Baumann. "An Investigation of Children's Ability to Comprehend Character Motives." *American Educational Research Journal*, vol. 25, 1988, pp. 441–462.

Shrout, Patrick E. and Joseph L. Fleiss. "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychological Bulletin*, vol. 86, no. 2, 1979, pp. 420–428.

Silverman, R., and J. D. Crandell. "Vocabulary Practices in Prekindergarten and Kindergarten Classrooms." *Reading Research Quarterly,* vol. 45, no. 3, 2010, pp. 318–340.

Snow, Catherine E., M. Susan Burns, and Peg Griffin. *Preventing Reading Difficulties in Young Children*. Washington, DC: The National Academies, 1998.

Song, S., M. Su, C. Kang, H. Liu, Y. Zhang, C. McBride-Chang, T. Tardif, H. Li, W. Liang, Z. Zhang, and H. Shu. "Tracing Children's Vocabulary Development from Preschool through the School-Age Years: An 8-Year Longitudinal Study." *Developmental Science*, vol. 18, 2015, pp. 119–131.

Spires, H. A., J. Gallini, and J. Riggsbee. "Effects of Schema-Based and Text Structure-Based Cues on Expository Prose Comprehension in Fourth Graders." *Journal of Experimental Education*, vol. 60, 1992, pp. 307–320.

Spörer, N., J. C. Brunstein, and U. Kieschke. "Improving Students' Reading Comprehension Skills: Effects of Strategy Instruction and Reciprocal Teaching." *Learning and Instruction*, vol. 19, 2009, pp. 272–286.

Sterbinsky, A., M. S. Ross, and D. Redfield. "Effects of Comprehensive School Reform on Student Achievement and School Change: A Longitudinal Study." *School Effectiveness and School Improvement*, vol. 17, 2006, pp. 367–397.

Stevens, R.J., N.A. Madden, R.E. Slavin, and A.M. Farnish. "Cooperative Integrated Reading and Composition: Two Field Experiments." *Reading Research Quarterly,* vol. 22, no. 4, 1987, pp. 433–454.

Sticht, Thomas G., Lawrence J. Beck, Robert N. Hauke, Glenn M. Kleiman, and James H. James. *Auding and Reading: A Developmental Model.* Alexandria, VA: Human Resources Research Organization, 1974.

Stringfield, S., A. Datnow, G. Borman, and L. Rachuba. *National Evaluation of Core Knowledge Sequence Implementation: Final Report.* Baltimore, MD: Center for Research on the Education of Students Placed at Risk, Johns Hopkins University and Howard University, 2000.

Sullivan, Daniel G. "A Note on the Estimation of Linear Regression Models with Heteroskedastic Measurement Errors." Chicago, IL: Federal Reserve Bank of Chicago, 2001.

Taboada, A., and J.T. Guthrie. "Contributions of Student Questioning and Prior Knowledge to Construction of Knowledge from Reading Information Text." *Journal of Literacy Research,* vol. 38, no. 1, 2006, pp. 1–35.

Taylor, B. M., P. D. Pearson, K. M. Clark, and S. Walpole. "Effective Schools and Accomplished Teachers: Lessons about Primary-Grade Reading Instruction in Low-Income Schools." *Elementary School Journal*, vol. 101, 2000, pp. 121–165.

Thurstone, L. L. *The Vectors of Mind.* Chicago: University of Chicago Press, 1935.

Thurstone L. L. *Multiple Factor Analysis*. Chicago: University of Chicago Press, 1947.

Tierney, R.J., and T. Shanahan. "Research on the Reading-Writing Relationship: Interactions, Transactions, and Outcomes." In *Handbook of Reading Research, vol. 2,* edited by R. Barr, M.L. Kamil, P. Mosenthal, and P.D. Pearson. Mahwah, NJ: Lawrence Erlbaum Associates, 1991, pp. 246–280.

Tompkins, V., Y. Guo, and L. M. Justice. "Inference Generation, Story Comprehension, and Language Skills in the Preschool Years." *Reading and Writing*, vol. 26, 2013, pp. 403–429.

Torgesen, Joseph K., Richard K. Wagner, Carol A. Rashotte, Elaine Rose, Patricia Lindamood, Tim Conway, and Cyndi Garvan. "Preventing Reading Failure in Young Children with Phonological Processing Disabilities: Group and Individual Responses to Instruction." *Journal of Educational Psychology*, vol. 91, no. 40, 1999, pp. 579–593.

Tourangeau, Karen, Christine Nord, Thanh Lê, Kathleen Wallner-Allen, Nancy Vaden-Kiernan, Lisa Blaker, and Michelle Najarian. *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011) User's Manual for the ECLS-K:2011 Kindergarten–Second Grade Data File and Electronic Codebook, Public Version* (NCES 2017-285). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, February 2017.

Trabasso, T., and M. Nickels. "The Development of Goal Plans of Action in the Narration of a Picture Story." *Discourse Processes,* vol. 15, 1992, pp. 249–275.

U.S. Department of Education, National Center for Education Statistics. *Early Childhood Longitudinal Study—Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for Kindergarten Through First Grade* (NCES 2002–05). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2002.

U.S. Department of Education, National Center for Education Statistics. "User's Manual for the ECLS–K Third Grade Public-Use Data File and Electronic Code Book." (NCES 2004–001). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2004.

U.S. Department of Education, National Center for Education Statistics. *The Nation's Report Card: 2015 Mathematics and Reading Assessments.* 2015. Available at http://www.nationsreportcard.gov/reading_math_2015/#?grade=4.

van den Broek, P. "The Causal Inference Maker: Towards a Process Model of Inference Generation in Text Comprehension." In D. A. Balota, G. B. Flores d'Arcais, and K. Rayner (Eds.), *Comprehension Processes in Reading*. New York: Routledge, 1990, pp. 423–446.

Velicer, W. F. "Determining the Number of Components from the Matrix of Partial Correlations." *Psychometrika*, vol. 41, no. 3, 1976, pp. 321–327.

Velicer, W. F., C. A. Eaton, and J. L. Fava. "Construct Explication through Factor or Component Analysis: A Review and Evaluation of Alternative Procedures for Determining the Number of Factors or Components." In *Problems and Solutions in Human Assessment*. Springer, US, 2000, pp. 41–71.

Vellutino, F.R., W.E. Tunmer, J.J. Jaccard, and R.S. Chen. "Components of Reading Ability: Multivariate Evidence for a Convergent Skills Model of Reading Development." *Scientific Studies of Reading*, vol. 11, no. 1, 2007, pp. 3–32.

White, H. L., Jr. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, vol. 48, 1980, pp. 817–838.

Whitehurst, G. J., F. L. Flaco, C. J. Lonigan, J. E. Fischel, and B. D. DeBaryshe, M. C. Valdez-Menchaca, and M. Caulfield. "Accelerating Language Development through Picture Book Reading." *Developmental Psychology*, vol. 24, 1988, pp. 552–559.

Wiig, Elisabeth, Wayne A. Secord, and Eleanor Semel. *Clinical Evaluation of Language Fundamentals Preschool–2 (CELF Preschool–2)*. San Antonio, TX: Pearson Assessments, 2004.

Williams, J., A.M. Nubla-Kung, S. Pollini, K.B. Stafford, A. Garcia, and A.E. Snyder. "Teaching Cause-Effect Text Structure Through Social Studies Content to At-Risk Second Graders." *Journal of Learning Disabilities*, vol. 40, no. 2, 2007, pp. 111–120.

Williams, J. P., K. M. Hall, K. D. Lauer, K. B. Stafford, L. A. DeSisto, and J. S. deCani. "Expository Text Comprehension in the Primary Grade Classroom." *Journal of Educational Psychology*, vol. 94, 2005, pp. 235–248.

Wittrock, M.C. "Learning as a Generative Process." *Educational Psychologist*, vol. 11, 1974, pp. 87–95.

Woodcock, Richard W., Kevin McGrew, and Nancy Mather. *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing, 2001.

Woodcock, Richard W., Kevin McGrew, and Nancy Mather. *Woodcock-Johnson III Normative Update*. Itasca, IL: Riverside Publishing, 2007.

Yoder, P. J., H. Spruytenburg, A. Edwards, and B. Davies. "Effect of Verbal Routine Contexts and Expansions on Gains in the Mean Length of Utterance in Children with Developmental Delays." *Language, Speech, and Hearing Services in Schools,* vol. 26, 1995, pp. 21–32.

Zucker, T. A., S. Q. Cabell, L. M. Justice, J. M. Pentimonti, and J. N. Kaderavek. "The Role of Frequent, Interactive Prekindergarten Shared Reading in the Longitudinal Development of Language and Literacy Skills." *Developmental Psychology,* vol. 49, no. 8, 2012, 1425–1439.

This page has been left blank for double-sided copying.

APPENDIX A: SUPPLEMENTARY INFORMATION ON THE SELECTION AND CHARACTERISTICS OF THE STUDY SAMPLE

This appendix describes the process by which the study obtained a sample of schools, classrooms, and students to assess relationships between instructional practices and student growth. We describe the selection of the initial study sample, the determination of the final analysis sample, the construction of analysis weights, and characteristics of the final analysis sample.

A. Selection of the study sample

Two key objectives guided the selection of the study sample. First, the study sample needed to be large enough to generate precise estimates of relationships between instructional practices and student growth. Second, the study sample needed to be selected in a way that maximized variation in student growth across classrooms and schools, because this variation was important for identifying instructional practices that were associated with greater student growth. This section describes the key steps by which we selected a study sample meeting these objectives (Figure A.1). These steps resulted in a large study sample, with the study's final analyses consisting of 10 districts, 83 schools, more than 1,000 classrooms, and nearly 5,000 students (Table A.1).

Figure A.1. Key steps in the sample selection process

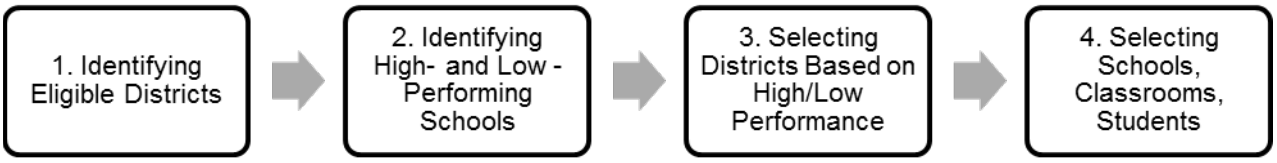


Table A.1. Number of districts, schools, classrooms, and students in the study

Type of sample member	Number of sample members	
	Initially selected for the study	In final analysis sample
Districts	10	10
Schools	141	83
Classrooms[a]	1,068	1,035
Prekindergarten and kindergarten	390	378
Grades 1 through 3	678	657
Students[b]	7,985	4,969
Prekindergarten and kindergarten	2,880	1,783
Grades 1 through 3	5,105	3,186

Source: Authors' calculations from study-collected sample information.
[a]Numbers of classrooms that were initially selected for the study are restricted to schools in the final analysis sample.
[b]Numbers of students who were initially selected for the study are restricted to classrooms in the final analysis sample.

1. Identifying eligible districts

The districts that were most suitable for this study were those that had large variation in student reading outcomes across schools. Variation in student outcomes was necessary for identifying instructional practices that could be related to such outcomes. If there were little to no variation in student outcomes, then estimates of relationships between instructional practices and student outcomes would always be close to zero.

We identified 44 districts (or geographically proximate clusters of districts) as eligible for the study because they met both of the following criteria:

- **Included a large number of Title I elementary schools.** These districts were likely to have both high- and low-performing schools that qualified for Title I. Accordingly, districts were eligible for the study only if they had 50 or more Title I elementary schools in the 2007–2008 school year.
- **Were located in states whose low-income students had relatively high average reading achievement.** Although the study needed both high- and low-performing Title I schools, we expected high-performing Title I schools to be harder to find. To maximize the opportunity for finding high-performing Title I schools, we focused on districts in states whose low-income students had relatively high 4th-grade reading assessment scores on the National Assessment of Educational Progress (NAEP). Specifically, those states needed to demonstrate NAEP reading proficiency results for students receiving free or reduced-price lunch that were equal to or above the national average for such students in 2007.

2. Identifying high- and low-performing schools in eligible districts

After identifying the 44 eligible districts, we examined, in greater detail, which of these districts were best suited to the study based on how many low- and high-performing Title I schools each district contained. Therefore, within each district, the key next step was to identify Title I elementary schools that were consistently low- or high-performing in reading achievement.

To measure school performance in reading, we obtained data on the percentage of 3rd-grade students in each elementary school who scored proficient or advanced on the state's reading assessment, referred to as the school's *proficiency rate*.[9] We used data from three years—the 2005–2006 through 2007–2008 school years—to obtain more reliable classifications of school performance than would be available using only one year of data.

With this data, we applied three criteria to identify consistently high- or low-performing schools. Each criterion identified different sets of schools. The criteria were as follows:

- **Criterion 1: Met a threshold based on the average proficiency rate across three years.** Under this definition, consistently high-performing schools were those with a three-year average proficiency rate equal to or greater than the median for Title I schools in the state. Consistently low-performing schools were those with an average proficiency rate under the

[9] The data came from schooldatadirect.org, a public-use database compiled by the Council of Chief State School Officers with support from the Bill & Melinda Gates Foundation. This database is no longer active.

25th percentile. The remaining schools—those with an average proficiency rate between the 25th and 50th percentiles—were considered medium-performing.

- **Criterion 2: Demonstrated an average proficiency rate that exceeded or fell short of expectations based on student disadvantage.** Because simple proficiency rates may reflect student background characteristics rather than schools' effectiveness, we adjusted each school's three-year average proficiency rate to account for its level of economic disadvantage. Specifically, we calculated the difference between a school's actual proficiency rate and its predicted proficiency rate based on the fraction of its students receiving free or reduced-price lunch. Schools with adjusted proficiency rates at the 80th percentile or above within their district were considered consistently high-performing; those at the 20th percentile or below were considered consistently low-performing. The remaining schools—those between the 20th and 80th percentiles—were considered to be meeting expectations.
- **Criterion 3: Met a threshold based on advanced proficiency (high-performing schools only).** For this categorization, we used the percentage of students scoring at the advanced (highest) level on the state's reading assessment, averaged across three years. We classified a school as consistently high-performing if this performance measure placed the school above the 75th percentile for Title I schools in the same state.

In consultation with an expert panel, we used a combination of the three criteria to arrive at a final classification of high- and low-performing schools in each district. The combination of criteria used depended on whether a state had a relatively lenient or stringent standard for deeming a student proficient (Bandeira de Mello et al. 2009). In states with a relatively stringent standard—a proficiency cutoff score that was effectively higher than the NAEP cutoff for scoring basic—schools needed to meet both criteria 1 and 2 to be considered consistently high-performing. In the remaining states, schools needed to meet all three criteria to be considered consistently high-performing; otherwise, if only a subset of criteria had been used, a large number of schools would have been (inappropriately) deemed high-performing due to the leniency of the state's proficiency standard. In all states, schools were classified as consistently low-performing if they met both criteria 1 and 2.

3. Selecting districts based on the availability of high- and low-performing schools

After identifying high- and low-performing schools in each district, we determined which of the eligible districts were most aligned with the study's goal of having a large study sample with substantial variation in student outcomes. From the set of 44 eligible districts, we sought to recruit into the study those districts that had larger numbers of consistently high- and low-performing schools and larger differences between the reading proficiency rates of high- and low-performing schools. We reached out to 18 districts, 10 of which agreed to participate in the study.

4. Selecting schools, classrooms, and students

In each of the 10 districts in the study, we selected schools, classrooms within those schools, and students within those classrooms to participate in the study.

Selection of schools. In each district, we identified schools that met three eligibility criteria: (1) classified as either consistently high-performing or consistently low-performing; (2) participated in Title I at the school-wide level, which meant that at least 40 percent of students in the school were eligible for free or reduced-price lunch;[10] and (3) had two or more classes in each of the grades from prekindergarten through grade 3. The study's target was to secure the participation of approximately five eligible high-performing schools and five eligible low-performing schools from each district. If more schools were eligible than needed, we randomly selected a subset of those schools.

Table A.2 lists the states in which the 10 participating school districts were located, the number of schools initially selected for the study from each state, and the number of schools included in the final analysis sample. Not all schools that were initially selected for the study agreed to participate, as described in Section B of this appendix. The 83 schools in the final analysis sample came from nine states that represented a diverse set of geographic regions.

Table A.2. Number of schools in the study, by state

	Number of schools	
State	Initially selected for the study	In final analysis sample
California	16	9
Florida and Ohio	16	8
Georgia	17	15
Massachusetts	13	9
New Mexico	13	8
New York	25	9
Tennessee	14	10
Texas	27	15
Total	**141**	**83**

Source: Authors' calculations from study-collected sample information

Selection of classrooms. To select classrooms for the study, we identified eligible classrooms in prekindergarten through grade 3 within each participating school. Eligible classrooms were general education classrooms in which a majority of total instruction and all language arts instruction were in English. As an exception, special education prekindergarten classrooms were also eligible for the study. In prekindergarten, one of the most common reasons why students are identified for special education services is a speech or language impairment, but such impairments may be temporary and get resolved by the time students enter elementary school. Therefore, because one of the key purposes of this study was to identify potentially promising ways to promote students' language development, we did not perceive a compelling reason to exclude special education prekindergarten classrooms.

At each grade level within each school, the study included up to three classrooms; if more than three were eligible, we randomly selected three to be in the study. The final analysis sample included 1,035 classrooms (Table A.1).

[10] When 40 percent of the students in a school are eligible for free or reduced-price lunch benefits, the school is eligible to use Title I funds for "school-wide" programs that serve all children in the school—that is, the school is not only Title I-eligible, but also eligible for school-wide Title I.

Selection of students. The study's target was to include five students per classroom. In each selected classroom, we obtained a list of enrolled students, from which we randomly selected eight students for the study plus three backup students. The assumption was that some of the initially selected students would drop out of the study due to lack of parental consent or transferring out of the school, leaving us with approximately five students per classroom. We included the backup students into the study only when the study team determined that this was needed due to low consent rates in a particular school and grade.

After initial selection of students, the study team determined which of the selected students were ineligible for the study. We deemed students ineligible if they had an individualized education program that indicated the student could not be assessed, or if the student would have required an alternative assessment, such as one that used Braille or involved sign language. In addition, from each group of siblings in the initially selected sample, the study team randomly selected only one student. This sampling procedure provided a final analysis sample of 4,969 students (Table A.1).

B. Determination of the final analysis sample

Most, but not all, of the schools, classrooms, and students that we initially selected for the study were included in the final analysis. Table A.3 summarizes the number of initially selected sample members, the number in the final analysis sample, and the key reasons sample members were dropped from the study.

As described earlier, the final analysis sample included 83 of the 141 initially selected schools. The other schools declined to participate in the study, or originally agreed to participate but then were not cooperative with data collection.

Within the participating schools, most of the classrooms (1,035 out of 1,068) initially selected for the study were included in the final analysis sample. Nevertheless, 33 classrooms (3 percent) were dropped from the study for various reasons. Across all grades, we dropped 20 classrooms by the spring of the study school year because they did not have any students with parental consent to participate in the study. This situation could have occurred because the teacher did not send home consent forms with the students, none of the parents gave their consent, all previously consenting students moved out of the classroom, or the classroom itself dissolved (with all students relocated to other classrooms). We dropped 13 classrooms because they were missing essential data; either the teacher declined to be observed, or none of the consenting students in the classroom completed both fall and spring assessments.

Table A.3. Determination of the final analysis sample

Group	Sample size
Schools	
Initially selected for the study	141
Left the study because:	
Declined to participate	44
Did not fully cooperate with data collection	14
In final analysis sample	83
Classrooms in prekindergarten and kindergarten[a]	
Initially selected for the study	390
Left the study because:	
Did not have any consenting students by the time of the spring tests	8
Teacher declined to be observed or no students had both fall and spring test scores	4
In final analysis sample	378
Classrooms in grades 1 through 3[a]	
Initially selected for the study	678
Left the study because:	
Did not have any consenting students by the time of the spring tests	12
Teacher declined to be observed or no students had both fall and spring test scores	9
In final analysis sample	657
Students in prekindergarten and kindergarten[b]	
Initially selected for the study	2,880
Left the study because:	
Did not have parental consent to participate	785
Had consent but no fall score on at least one assessment	74
Had fall score but no spring score on any of the fall assessments	238
In final analysis sample	1,783
Students in grades 1 through 3[b]	
Initially selected for the study	5,105
Left the study because:	
Did not have parental consent to participate	1,498
Had consent but no fall score on at least one assessment	92
Had fall score but no spring score on any of the fall assessments	329
In final analysis sample	3,186

Source: Authors' calculations from study-collected sample information and fall and spring tests administered by the study team.

[a]Numbers of classrooms are restricted to schools in the final analysis sample.

[b]Numbers of students are restricted to classrooms in the final analysis sample.

Of the students we initially selected for the study, slightly more than 60 percent (4,969 out of 7,985 students) contributed to the final analysis. The primary reason we dropped students from the study was that they did not receive parental consent to be administered assessments. Other students were dropped because we were not able to conduct a fall or spring assessment (for example, because they moved to a different school).

C. Construction of analysis weights

When assessing relationships between instructional practices and student growth, this study sought to generate findings that were applicable to all students who were eligible for the study within the 10 study districts. As discussed in Section A of this appendix, the eligible study population consisted of students who were able to take assessments, enrolled in general education classrooms with most instruction in English, and attending either high- or low-performing elementary schools with Title I schoolwide programs. However, as described earlier, not all eligible schools, classrooms, and students were selected for the study, and not all those we initially selected for the study contributed to the final analysis.

We constructed and used analysis weights to help ensure that the final analysis sample would be representative of the eligible study population. The key threat to having a representative sample was that some groups of eligible schools, classrooms, or students had a greater likelihood of being included in the final analysis sample than others. Without use of analysis weights, those groups with a greater likelihood of inclusion would be overrepresented in the analysis sample. For example, because the study generally selected about equal numbers of students from each classroom, students in smaller classrooms had a greater probability of being selected; without weights to account for this design, students from smaller classrooms would be overrepresented in the final analysis sample. Likewise, if classrooms in certain schools were more likely to cooperate with data collection than those in other schools, classrooms from the more cooperative schools would be overrepresented unless we used weights to account for these differences in response rates.

We prepared weights for schools, classrooms, and students in the final analysis sample. At each level, the weight was inversely proportional to the probability of being selected into the sample and the participation rate among those who were selected. Therefore, sample members with lower probabilities of being selected or those that were less likely to participate—that is, the sample members who would otherwise be underrepresented—had larger weights. We calculated these selection probabilities and participation rates separately in specific groups of schools, classrooms, and students to reflect the potential for these groups to differ in their likelihood of being selected or of providing data after being selected. When defining groups for calculating these selection probabilities and participation rates, we grouped sample members at each level by the stratum from which they were selected. We provide details next.

- **To calculate school analysis weights**, we grouped eligible schools by the combination of district and performance level (high or low). Within each group, we first calculated the probability of being selected for the study. Because the study typically selected about five schools per group, the probability of selection was generally lower in groups with more schools. Second, among selected schools that were eligible, we calculated the study participation rate. We multiplied the selection probability and the participation rate and took the reciprocal to generate the final school analysis weight for participating schools.
- **To calculate classroom analysis weights**, we grouped classrooms by the combination of school and grade. Within each group defined by school and grade, we first calculated the probability of being selected for the study (given that the classroom's school participated in the study). Because the study selected one classroom per teacher and up to three classrooms per group, the probability of selection was lower for classrooms taught by teachers who also

taught other eligible classrooms, and for classrooms in larger groups. Second, among classrooms selected for the study, we calculated the study participation rate—the fraction of classrooms with at least one observation session conducted. We multiplied the selection probability and the participation rate, took the reciprocal, and multiplied by the school's final analysis weight to generate the final classroom analysis weight for participating classrooms.

- **To calculate student analysis weights,** we grouped students by classroom. Within each classroom, we first calculated the probability of being selected for the study (given that the student's classroom had at least one observation session). Because the study typically selected similar numbers of students per classroom and no more than one student per family, the probability of selection was lower for students in larger classrooms and for students with siblings who were also selected for the study. Second, among selected students who were eligible, we calculated the study participation rate—the fraction who had parental consent to participate and completed both the fall and spring assessments. We multiplied the selection probability and the participation rate, took the reciprocal, and multiplied by the classroom's final analysis weight to generate the final student analysis weight for participating students.

D. Characteristics of the final analysis sample

The study's findings on the relationships between instructional practices and student growth pertain to the types of schools, teachers, and students in the study. Here, we describe the background characteristics of the study participants and the performance of students on the study assessments to provide insight on the populations to which the study's results may be most relevant.

1. School characteristics

Schools in the study had high concentrations of low-income and minority students, even compared with other Title I elementary schools in the United States. On average, four of every five students in the study schools received free or reduced-priced lunch compared with three of every five students in all Title I schools (Table A.4). The study schools were also 94 percent nonwhite compared with 50 percent nonwhite in all Title I schools.

The differences between study schools and Title I schools nationwide resulted from our strategy for selecting study districts. As discussed earlier, we recruited particularly large districts into the study because those districts had large numbers of high-performing and low-performing Title I schools. Because poverty, minority status, and large school size are more prevalent in large districts, those characteristics were also more prevalent in our study sample.

Table A.4. Characteristics of schools in the study and all U.S. Title I elementary schools

Characteristic	Average for study schools	Average for all Title I elementary schools in U.S.
Percentage of students receiving free or reduced-price lunch	81	59
Race/ethnicity (percentage of students)		
Asian or Pacific Islander	10	4
Black, non-Hispanic	38	16
Hispanic	44	26
White, non-Hispanic	6	50
Other	1	4
School size (number of students)	618	446
Number of schools—range[a]	**80–83**	**39,503–41,663**

Source: Common Core of Data, 2011–2012.

Note: Elementary schools are defined as those whose lowest grade is grade 3 or below and whose highest grade is grade 8 or below. Title I status is not available for schools in Georgia, so the final column excludes Georgia.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

2. Teacher characteristics

Most teachers in the study had high levels of teaching credentials as measured by advanced degrees, certification, and teaching experience. More than half of the teachers had a master's degree or above, and nearly all (95 percent or more) were fully certified to teach, having completed all certification requirements (Table A.5). About 85 percent of teachers in the study had 5 or more years of teaching experience, and, on average, the teachers had about 15 years of experience.

3. Student characteristics

Of the various characteristics of students in the study, the students' performance on the study assessments was potentially most informative about the types of populations to which the study's findings might be relevant. For example, instructional practices that promote the growth of low-performing students may not necessarily do so for high-performing students. In what follows, we describe the performance of students in the study by expressing their test scores as percentiles within the national population. For each assessment, we report the national percentile of study students who outperformed 20, 50, and 80 percent of other students in the study—referred to as low, middle, and high performers. For the basic language skills assessment, national norms were not available for the total language skills score (combining the four subtests that we administered in the study), so we report percentiles based on two indices—one measuring receptive language, and one measuring semantic knowledge—that use certain combinations of those subtests.

Table A.5. Characteristics of teachers in the study (percentages unless otherwise noted)

Characteristic	Prekindergarten and kindergarten	Grades 1 through 3
Gender		
Female	97	93
Male	3	7
Race/ethnicity		
Black, non-Hispanic	42	41
Hispanic	15	17
White, non-Hispanic	37	37
Other	9	6
Highest degree attained		
Bachelor's or below	42	39
Master's or above	58	61
Certification status		
None, temporary, or provisional	5	4
Full	95	96
Teaching experience		
Fewer than 5 years	14	13
5 to 15 years	45	52
More than 15 years	41	35
Average years teaching	15	14
Number of teachers—range[a]	**326–339**	**565–577**

Source: Teacher survey administered by the study team.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

Students in the study were low-performing compared with the national population. If the test scores of students in the study resembled those of the national population, the national percentiles of low, middle, and high performers in the study would be 20, 50, and 80. Instead, their national percentiles were all lower than those benchmarks (Table A.6).

On average, prekindergarten and kindergarten students in the study underperformed students of the same age nationwide. They had especially low scores on both of the language skills indices; for example, middle performers in the study scored at only the 10th to 16th percentiles nationwide on those indices. Prekindergarten and kindergarten students in the study generally earned higher national percentiles on background knowledge and listening comprehension, but were still well below their national peers, with middle performers scoring at the 25th to 37th percentiles on those assessments. On most assessments, study students generally earned higher national percentiles in the spring than fall and, at both points in time, demonstrated wide variation in achievement across low, middle, and high performers.

Study students in grades 1 through 3 also underperformed their peers nationwide, but not by as much as the study students in prekindergarten and kindergarten did. They performed most poorly on the receptive language portion of the language skills assessment and the background knowledge assessment, with middle performers scoring at no more than the 30th national percentile. The performance of the study students was somewhat higher for semantic knowledge, listening comprehension, and reading comprehension; national percentiles for middle performers ranged from the 34th to 45th percentiles.

Table A.6. Test performance of students in the study, expressed as percentiles in the national population

Student outcome	Percentile in the national population			Number of students
	Low performer among students in the study	Middle performer among students in the study	High performer among students in the study	
Prekindergarten and kindergarten				
Language skills: receptive language[a]				
Fall	2	10	37	992
Spring	4	16	50	992
Language skills: semantic knowledge[a]				
Fall	2	13	37	992
Spring	4	16	50	992
Background knowledge[b]				
Fall	5	26	57	1,030
Spring	7	25	54	1,030
Listening comprehension				
Fall	4	25	61	1,716
Spring	9	37	73	1,716
Grades 1 through 3				
Language skills: receptive language				
Fall	5	30	58	3,173
Spring	8	30	58	3,173
Language skills: semantic knowledge[c]				
Fall	7	34	70	2,066
Spring	12	39	75	2,066
Background knowledge				
Fall	7	26	60	987
Spring	8	27	58	987
Listening comprehension				
Fall	13	39	68	3,176
Spring	18	45	73	3,176
Reading comprehension[d]				
Spring	15	44	73	1,020

Source: Authors' calculations using data from the fall and spring tests administered by the study team.

Note: Low, middle, and high performers are defined as those who scored at the 20th, 50th, and 80th percentiles among students in the study.

Table reads: On the fall receptive language assessment in prekindergarten and kindergarten, a middle performer—that is, a student who scored at the 50th percentile among students in the study—scored at the 10th percentile among all students nationwide.

[a]National norms were available beginning at age 5, so prekindergarten students were excluded from this analysis.

[b]National norms were available beginning in kindergarten, so prekindergarten students were excluded from this analysis.

[c]National norms were available up through age 8, so grade 3 students were excluded from this analysis.

[d]National norms were available only for spring of grade 3.

Aside from low performance, another distinguishing characteristic of the study sample was the high prevalence of non-English home languages. Thirty percent of prekindergarten and kindergarten students in the study spoke a language other than English at home (Table A.7), compared with 16 percent of kindergarteners nationwide (Aud et al. 2013). Some of the students who spoke a language other than English at home received English as a Second Language (ESL) services; those services reached 11 percent of all study students in the lower grades and 17 percent in the upper grades.

Table A.7. Characteristics of students in the study (percentages)

Characteristic	Prekindergarten and kindergarten	Grades 1 through 3
Gender		
Female	49	51
Male	51	49
Overage for grade	1	4
Home language		
English	69	70
Spanish	22	24
Other	8	7
Receives English as a Second Language services	11	17
Receives special education services	9	8
Number of students—range[a]	**1,550–1,783**	**2,673–3,186**

Source: Authors' calculations using study-collected sample information and teacher student reports.

[a]Sample sizes are presented as a range, based on the data available for each row in the table.

APPENDIX B: SUPPLEMENTARY INFORMATION ON STUDY INSTRUMENTS AND DATA COLLECTION

To examine the relationships between instructional practices and student growth in language and comprehension, the study team collected data on (1) instructional practices by conducting classroom observations and (2) student growth by administering assessments to students. In addition, to obtain background information on the study sample, the study team surveyed teachers about their professional characteristics and the characteristics of their students. This appendix provides details on the instruments and data collection procedures used to obtain information on instructional practices (Section A), student growth (Section B), and background characteristics (Section C). All data described in this appendix were collected in the 2011–2012 school year.

A. Measuring instructional practices

In this section, we first describe the classroom observation instrument used in this study, then briefly outline the method we used to collect the observation data, and finally report on the degree of consistency among observers who rated the same instruction.

1. Observation instrument

To capture reliable information about the instructional practices that teachers in the study used in their classrooms, the study team developed a new instrument called the Observation of Language and Literacy Instruction (OLLI). As discussed in Chapter II of the main report, the study team developed the OLLI by conducting an extensive literature review of the aspects of instruction that prior research suggested might influence language development and comprehension. The OLLI included a large number of items—285 in total—to measure a comprehensive set of these aspects of instruction, including items that were based on competing theories of teaching.

The items on the OLLI covered ten broad dimensions of instruction (Table B.1). Four dimensions—classroom context, classroom climate, time management, and student engagement—covered general aspects of teaching that could promote effective instruction. Another four dimensions—language use, higher-order thinking, world knowledge, and vocabulary (outside of reading)—covered aspects of instruction that could support students' language development. The remaining two dimensions—book or text sharing and reading comprehension strategies—covered aspects of instruction related specifically to literacy.

Each dimension of the OLLI was informed by particular strands of literature on instructional practices:

- The *language use* dimension of the OLLI was informed by the literature on practices that expose students to rich language models and practices that extend and elaborate students' own language (Berko Gleason 2005; Baker et al. 2006; Gersten et al. 2005; Ruddell 1978; Stevens et al. 1987; Taboada and Guthrie 2006; Murphy et al. 2009; Spörer et al. 2009).
- The *higher-order thinking* dimension of the OLLI was informed by the literature on practices that help students operate at a high cognitive level during interpretive experiences

(Correnti and Rowan 2007; Wittrock 1974; Andre 1979; Taboada and Guthrie 2006; Taylor et al. 2000).

- The *world knowledge* dimension of the OLLI was informed by the literature on practices that help students develop background knowledge for reading a wide variety of content and texts (Nagy et al. 1987; Beck and McKeown 1991; Durso and Coggins 1991).
- The *vocabulary* dimension of the OLLI was informed by the literature on practices that help students develop rich, multifaceted definitions of words and that engage students actively in trying to use the words in a meaningful context (Blachowicz and Fisher 2007; Carlisle and Rice 2002; NICHD 2000; Pressley 2000; Beck and McKeown 1991).
- The *book or text sharing* dimension of the OLLI was informed by the literature on practices that help students form coherent mental representations of text (Gagne and Memory 1978; Neuman 1988; Spires et al. 1992; Koskinen et al. 1989; Law 2008; Stevens et al. 1987; NICHD 2000; Williams et al. 2005; Williams et al. 2007; Marley et al. 2007; Casteel 1993; Goldman and Varnhagen 1986; Shannon et al. 1988; Trabasso and Nickels 1992; van den Broek 1990; McKeown et al. 2009).
- The *reading comprehension strategies* dimension of the OLLI was informed by the literature on practices that help students learn mental operations for processing text in particular ways (Brown et al. 1996; Chan and Cole 1986; Duffy et al. 1986; Eilers and Pinkley 2006; Kelly et al. 1994; NICHD 2000; Rosenshine et al. 1996; Spörer et al. 2009; Williams et al. 2005; Williams et al. 2007).

Four other dimensions of the OLLI that captured general instructional practices—classroom context, classroom climate, time management, and student engagement—were developed by borrowing or adapting items from other commonly used observation instruments, including the Classroom Assessment Scoring System (CLASS; Pianta et al. 2006) and Teacher Behavior Rating Scale (TBRS; Landry et al. 2001).

As noted in Chapter II, the sections on text-related and vocabulary instruction included items to capture whether activities occurred as part of pre-reading, during-reading, or post-reading instruction. These distinctions applied regardless of the subject being taught. Specifically, whenever a teacher engaged students in discussing a text that they were about to read (for English language arts [ELA], mathematics, social studies, or science), the activity was coded as pre-reading. Whenever a teacher engaged the students in reading a text in class, the activity was coded as occurring during reading; and whenever a teacher engaged students in discussing a text that they had just read (that same day), the activity was coded as post-reading. In order to help observers distinguish between these phases, especially during non-ELA lessons, we provided extra practice during training, and included specific video exemplars to help illustrate the coding rule. For example, it was easy to know to code an activity as post-reading if the observer saw students reading a text and the teacher following up with a discussion of the text. And it was easy to know not to code an activity as post-reading if a teacher was discussing texts read on previous days. The more difficult decision was when the observers did not witness the students reading a text, but it was not clear from the discussion whether they did so earlier in the same day; in these situations, the discussion was not coded as a post-reading activity.

Table B.1. Dimensions of instruction included in the Observation of Language and Literacy Instruction

Dimension	Theory	Focus of items	Number of items
General aspects of instruction			
Classroom context	A low child-to-adult ratio and variation in classroom structures can support student learning.	Number of children and adults present; classroom structure (whole group, small group, partners); types of activities occurring	35
Classroom climate	Students learn best when they feel safe and cared for.	Incidence of positive and negative interactions	33
Time management	Students learn when they are in classrooms that are well-managed, with little downtime.	Amount of teaching time lost due to disruptions, transitions, and distractions	3
Student engagement	Students who are actively engaged in learning are likely to learn more.	Teacher enthusiasm; variation in activities; numbers of students called on or spoken to; encouragement to participate	23
Supports for language development			
Language use	Children learn language by observing and using language.	Amount and purposes of teachers' talk; clarity and grammatical accuracy of teachers' speech; efforts to expand students' language through open-ended questions and in-depth conversations	17
Higher-order thinking	Children who are engaged in inferencing, logical analysis, and evaluation will be better prepared to do so within reading.	Frequency and extent (time, number of questions) of higher-order reasoning; time allowed for students to respond to higher-order questions	7
World knowledge	Children need information about the natural and social worlds to interpret and learn new information.	Amount of information taught; efforts to engage students with the information; approaches used to present the information; relating world knowledge to themes or prior knowledge	24
Vocabulary (outside of reading)	Vocabulary knowledge—a bridge between world knowledge and language—is essential for reading and writing success.	Frequency and extent of vocabulary instruction; how words were explained (definition, synonym, illustration) and use of multiple approaches to explanation	11
Literacy-focused instruction			
Book or text sharing	Students should be well-practiced in comprehending and interpreting the meaning of texts.	Context for reading (amount of guided reading or listening, oral or silent reading, types of books read); activities that focused on text meaning (previewing, questioning, retelling, prior knowledge connections, teacher feedback); vocabulary instruction during reading	125
Reading comprehension strategies	Teaching students intentional ways of thinking during reading can improve comprehension.	Amount of strategy teaching (summarizing, questioning, visualizing, monitoring); emphasis on what, when, and why of strategy use	7

Source: Authors' compilation.

The OLLI included three basic types of items: (1) occurrence, (2) intensity, and (3) quality. Some items recorded the basic occurrence of an action, such as whether or not the teacher talked about the characters in a book. Other items recorded the intensity of an action or amount of a practice, such as how many words were defined during a vocabulary lesson. Still other items

focused on the quality of an action, such as the degree to which post-reading discussion was focused on content and was coherent.

Decisions about the types of items used—occurrence, intensity, or quality—were informed, in part, by the literature review described earlier. For example, the research studies referenced earlier that informed the language use dimension suggested that (1) classrooms vary in the amount of time teachers spend speaking with students and the clarity and correctness of their language, (2) instructional techniques to extend students' language can help promote their language growth, and (3) teachers' speech intended for social or instructional purposes tends to model richer language than speech intended only to give instructions or manage behavior. Accordingly, we included (1) an intensity item about how much time teachers spoke with students and a quality item about the clarity and correctness of their speech, (2) a set of occurrence items about the types of techniques teachers used to encourage student language, and (3) a quality item about the main purpose of teachers' talk.

2. Method for collecting observation data

In the spring of 2012, the study team recruited and trained observers to conduct the classroom observations. Approximately 100 trainees (80 percent with classroom experience as either teachers or teacher's aides, and 100 percent with undergraduate degrees) underwent a 10-day training session. This training included receiving lessons from experts on each of the components of instruction on the OLLI, viewing exemplars of practice (via video recordings), and practicing using the OLLI to rate video recordings of classroom instruction in prekindergarten through grade 3 for each dimension of the OLLI. In addition, the training included two days of practice applying the full OLLI (not just individual dimensions) to rate video recordings of classroom instruction. Training was conducted by senior survey researchers with extensive experience in conducting classroom observations and conducting trainings similar to this one.

At the end of the training, the trainees had to pass a two-part check of reliability to be certified to conduct observations in the study classrooms. First, they conducted a live observation in a classroom, accompanied by a trainer who served as the gold standard for the OLLI ratings. This observation was a full session (consisting of six 15-minute segments of instruction, as discussed in more detail below). Second, they viewed and rated video recordings of three teachers, covering six 15-minute segments of instruction per teacher. These videos had been previously coded by the training team. To be certified, each observer had to agree exactly with the gold standard (trainers') ratings on 80 percent of the items on the OLLI, and at least 75 percent of the items within each dimension of the OLLI. Of the 100 trainees, 92 were certified. Of the 92 certified observers, 81 conducted observations for the study; the remaining observers dropped out for a variety of personal reasons (including illness, scheduling conflicts, and securing full-time employment).

In the spring of 2012, the certified observers conducted observations in 1,041 study classrooms, of which 1,035 had all other necessary data in the study to be in the final analysis sample (see Appendix A). The study's goal was to conduct four observations per classroom. This occurred in 94 percent of the classrooms, and nearly all of the remaining classrooms had three observations (Table B.2). Each observation in a classroom was conducted by a different observer on a different day so that observer effects—biases or errors by individual observers—could

offset each other when averaged across multiple observations (see Appendix C for a further discussion of observer effects).

Table B.2. Classrooms in the study with specified numbers of observation sessions (percentages)

Number of observation sessions conducted	Percentage of classrooms
One or two observation sessions	1
Three observation sessions	5
Four observation sessions	94
Number of classrooms	**1,035**

Source: Authors' calculations from classroom observations conducted by the study team.

Because teachers' practices may vary by subject area, we planned to conduct half of the observations in the morning, when literacy instruction was most likely to occur, and half in the afternoon, when content-area (science and social studies) instruction was most likely to occur. Accordingly, most classrooms (68 percent) had equal numbers of morning and afternoon sessions—usually two each (Table B.3). In 26 percent of the classrooms, we conducted more morning than afternoon sessions, and in 6 percent more afternoon than morning.

Table B.3. Classrooms in the study with specified proportions of morning and afternoon observation sessions (percentages)

Time of day of observation sessions	Percentage of classrooms
More morning than afternoon observation sessions	26
Equal numbers of morning and afternoon observation sessions	68
Fewer morning than afternoon observation sessions	6
Number of classrooms	**1,035**

Source: Authors' calculations from classroom observations conducted by the study team.

Each observation session was approximately two hours long and was divided into six 20-minute segments (Table B.4). During each segment, observers focused on the classroom for 15 minutes, taking notes as needed. Then, after the 15 minutes elapsed, they spent 5 minutes rating the segment using the OLLI. They repeated this sequence until the two-hour period ended.

Table B.4. Structure of an Observation Session

Segment	Observation time (minutes)	Rating time using OLLI (minutes)
1	15	5
2	15	5
3	15	5
4	15	5
5	15	5
6	15	5
Total	**90**	**30**

3. Consistency between observers

During the observation data collection period, we monitored interrater reliability—the level of agreement among observers. Although each observation session usually had only one observer, the study assigned multiple observers to some of the observation sessions to check for interrater reliability. All of the 81 certified observers who conducted observations for the study co-observed one observation session (six segments of instruction) with another observer. Using data from those sessions, we calculated the exact agreement rate—the percentage of item scores in which observers in the same session came to exact agreement. The exact agreement rate ranged from 72 to 93 percent across dimensions of the OLLI, for an average of 83 percent (Table B.5).

As discussed in Chapter II and Appendix C, the items of the OLLI were ultimately grouped into 13 summary measures of instructional practices. Appendix C provides information on the interrater reliability of the summary measures.

Table B.5. Rate of agreement between observers who rated the same observation session

Dimension of observation instrument	Rate of exact agreement (percentage)
Classroom context	93
Classroom climate	89
Time management	72
Student engagement	84
Language use	82
Higher-order thinking	73
World knowledge	85
Vocabulary (outside of reading)	78
Book or text sharing	88
Reading comprehension strategies	85
Average across dimensions	**83**
Number of observation sessions	**42**
Number of observers	**81**

Source: Authors' calculations from classroom observation data.

Note: The rate of exact agreement is the percentage of item scores in which observers in the same observation session were in exact agreement. For eight sessions in which only a trainer was available to be paired with a regular observer, ratings assigned by both the trainer and regular observer were used.

B. Measuring student growth

We administered assessments to students in the study in both fall 2011 and spring 2012 to measure their growth in language and comprehension. The assessments measured students' basic language skills, background knowledge in science and social studies, listening comprehension, and reading comprehension (Table B.6).

Table B.6. Student assessments administered in the study

Domain of language and comprehension	Name of assessment	Grades
Basic language skills	Clinical Evaluation of Language Fundamentals Preschool–Second Edition[a] (Receptive Language Index)	PK–K
	Clinical Evaluation of Language Fundamentals–Fourth Edition[b] (Receptive Language Index)	1–3
Background knowledge	Early Childhood Longitudinal Study–Kindergarten Class of 1998–99 General Knowledge Assessment[b]	PK–1
Listening comprehension	Woodcock-Johnson III Tests of Achievement, Oral Comprehension Subtest[c]	PK–3
Reading comprehension	Early Childhood Longitudinal Study–Kindergarten Class of 1998–99 Third Grade Reading Assessment[d]	2–3

Source: Authors' compilation.

[a] Wiig et al. (2004).

[b] Semel et al. (2003).

[b] U.S. Department of Education (2002).

[c] Woodcock et al. (2001, 2007).

[d] U.S. Department of Education (2004); Pollack et al. (2005).

K = kindergarten; PK = prekindergarten.

We chose these assessments because they (1) covered the key domains of language and comprehension that the study sought to measure; (2) had evidence of being valid (measuring the knowledge or skills that were intended to be measured) and reliable (producing consistent scores for the same individual in the same circumstances); (3) could differentiate students with different skill levels, even within a generally low-achieving population; and (4) were used in prior research on students with age and socioeconomic status similar to those in this study.

In the remainder of this section, we first provide more detail on the domains of language and comprehension covered by each assessment and describe how these assessments were administered. We then specify the types of scores obtained from the assessments, summarize evidence on the assessments' reliability and validity from test publishers' information, and describe the degree to which scores varied reliably across students in our study sample.

1. Domains of language and comprehension assessed by the study

Basic language skills. Basic language skills encompass a range of skills and abilities, from understanding and recognizing the smallest units of sound in language to using a variety of words correctly in a social context. Experts in children's language development distinguish among these critical skills and abilities, which include phonology (how sounds operate), morphology (how words are formed from smaller units of meaning), syntax (grammar), semantics (word meaning), and pragmatics (use of language in a social context) (Brassard and Boehm 2007; Snow et al. 1998). Measures of basic language skills in young children are strongly related to subsequent reading comprehension in elementary school (National Early Literacy Panel 2008).

We used two assessments to examine students' basic language skills: the Clinical Evaluation of Language Fundamentals Preschool–Second Edition (CELF P–2) for prekindergarten and

kindergarten and the Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF–4) for grades 1, 2, and 3. The CELF P–2 was designed as a downward extension of the CELF–4.

The CELF P–2 and CELF–4 originally contain 11 and 18 subtests, respectively, and the study selected four subtests from each assessment to administer to the participating students. We selected these subtests because they (1) captured multiple dimensions of language skills; (2) were available in both versions of the CELF, allowing us to measure the same dimensions of language skills across the age span of the study, and (3) were associated with reading comprehension in past studies. The four subtests were:

- **Concepts and Following Directions** required students to point to pictures in response to oral commands. It was designed to assess receptive language (listening comprehension), syntax, working memory, and understanding of basic concepts. Composite measures that include this subtest have demonstrated moderate to high correlations with reading comprehension (Catts et al. 2008; Jarmulowicz et al. 2008; Scott et al. 2008).
- **Expressive Vocabulary** required students to name pictures of people, objects, or activities. It was designed to assess semantics and expressive vocabulary—the words that students are able to use to convey thoughts. Expressive vocabulary is commonly used in research studies to measure oral language skills and background knowledge, so including such a measure helped to relate student language achievement in this study to other research. Higher achievement on early expressive vocabulary has a consistent, moderate-sized relationship with later reading comprehension across a large number of studies, with large numbers of children (National Early Literacy Panel 2008). Justice et al. (2010) found that a higher degree of implementation of a literacy intervention was associated with an improvement in CELF P–2 Expressive Vocabulary subtest scores.
- **Word Classes** required students to identify or describe the relationship between two related words, such as whole-part, spatial, and temporal relationships. It was designed to assess semantics, receptive language, and expressive vocabulary. The subtest had two forms, one for ages 4 to 7 (Word Classes I) and another for ages 8 to 21 (Word Classes II). Prior research has included this subtest within a receptive language index that was associated with reading comprehension in third grade (Jarmulowicz et al. 2008).
- **Sentence Structure** required students to point to an illustration that represented a given sentence. It was designed to assess syntax, morphology, and receptive language. Measures of syntax generally are among the early language measures that are most closely related to later reading comprehension (National Early Literacy Panel 2008). This subtest was part of the receptive language index that Jarmulowicz et al. (2008) found to be associated with reading comprehension. Glenn-Applegate et al. (2010) found associations between the CELF P–2 Sentence Structure and narrative skills, which in turn were associated with reading comprehension. Justice et al. (2010) found that stronger implementation of a literacy intervention was associated with improvement in CELF P–2 Sentence Structure scores. This subtest from earlier versions of the CELF has also been part of composite scores associated with reading comprehension (Catts et al. 2008; Torgesen et al. 1999).

Background knowledge. Students' background knowledge includes their familiarity with basic concepts (such as space and time) and with the social, physical, and biological world.

Reading experts and others believe that background knowledge helps students extract meaning from texts (Hirsch 2003, 2006; Hoover and Gough 1990). For example, an understanding of time enables students to sequence events in a story. Background knowledge can also help students understand the context of the words they read, beyond simply understanding the words' literal definitions (Snow et al. 1998). In prior research, background knowledge of social studies and science content in kindergarten has been positively associated with reading achievement in grades 1, 3, and 5 (Claessens et al. 2009; Duncan et al. 2007).

We assessed the background knowledge of students in prekindergarten, kindergarten, and grade 1 with the Early Childhood Longitudinal Study–Kindergarten Class of 1998–99 (ECLS–K) General Knowledge Assessment (U.S. Department of Education 2002). This measure included assessment of both science (including earth and space, life, and physical sciences) and social studies (including culture, history, geography, government, and economics). During the assessment, students were shown pictures related to science and social studies, and they were asked to orally name the picture, describe what it means, or, for some items that contained four pictures, point to the correct answer.

We did not assess background knowledge in grades 2 and 3 for several reasons. The ECLS–K did not include a grade 2 measure, and the grade 3 background knowledge measure in ECLS–K was devoted solely to science. Furthermore, the study's priority in grades 2 and 3 was to assess reading comprehension, a domain that required students to apply their background knowledge.

Comprehension (listening and reading). Comprehension is the understanding of language that is spoken (listening comprehension) or written (reading comprehension). The close connection between listening comprehension and reading comprehension has long been recognized and demonstrated empirically. A meta-analysis of 30 independent studies indicates a relationship between kindergarten listening comprehension and later reading comprehension through age 7 (National Early Literacy Panel 2008). Additional studies demonstrate that the correlation between listening comprehension and reading comprehension persists well beyond these age levels (Sticht et al. 1974; Vellutino et al. 2007).

Preschoolers can comprehend text better that is read aloud to them than text that they read themselves (Carlisle and Rice 2002). As students gain word reading skills and fluency by grades 2 and 3, it becomes possible to assess their reading comprehension directly (Keenan et al. 2008). For this reason, we assessed listening comprehension in all grades in the study and reading comprehension in grades 2 and 3, as described in further detail below. We did not measure reading comprehension for students in grade 1 (or earlier) because results are often misleading, with first-grade measures of reading comprehension typically aligned too closely with decoding or word reading skills to represent a truly independent measure of reading comprehension (Francis et al. 2005; Keenan et al. 2008; Nation and Snowling 1997).

We assessed ***listening comprehension*** with the Woodcock-Johnson III (W-J III) Tests of Achievement, Oral Comprehension subtest (Woodcock et al. 2001, 2007) in all of the study grades from prekindergarten through grade 3. The W-J III Oral Comprehension subtest asked students to verbally supply the missing key word that completed an oral passage.

We assessed the ***reading comprehension*** of students in grades 2 and 3 with the ECLS–K Third Grade Reading Assessment (U.S. Department of Education 2004; Pollack et al. 2005). The content of this assessment was adapted from the 1992 and 1994 National Assessment for Educational Progress (NAEP) Reading Frameworks and included four types of reading comprehension skills: (1) identifying the main point of a passage, (2) developing interpretation, (3) connecting text to background knowledge, and (4) evaluating text objectively. During the assessment, students read passages and responded orally to questions that an assessor asked aloud. Although most items focused on reading comprehension, additional items assessed basic skills (such as recognition of letters and decoding) and vocabulary to provide information on students performing at lower levels.

Because the ECLS–K Third Grade Reading Assessment was originally designed for students in grade 3, we also needed to ensure that there were sufficient numbers of items appropriate for students in grade 2. Therefore, we included additional passages and reading comprehension questions from the Early Childhood Longitudinal Study–Kindergarten Class of 2010–11 Second Grade Reading Assessment (Tourangeau et al. 2017). These items covered similar skills as, but at a somewhat less advanced level than, the third-grade items.

2. Administration of the assessments

The assessments were administered with three key features: (1) administration to students was one-on-one, (2) study team members who administered the assessments received extensive training, and (3) the difficulty of the assessment items adapted to the students' ability levels.

One-on-one administration. Trained assessors from the study team administered the assessments to each student individually by computer. This approach allowed the assessment process to be sensitive to the needs of the young children in the study. For example, the assessors could proceed at a pace that was suited to each individual student and could provide encouragement to stay on task.

On average, the amount of time needed for students to complete the full battery of assessments was 55 to 60 in the lower grades (prekindergarten through grade 1) and 80 to 85 minutes in the upper grades (grades 2 and 3). For students in the lower grades, the administration of the assessments occurred in one session. For students in the upper grades, the 40-minute reading comprehension assessment was administered on a second day to minimize student burden and fatigue.

Training and monitoring of assessors. Seventy-five field assessors and 12 field team leaders received extensive training before administering the assessments. The training, led by senior survey researchers with extensive experience in administering the measures and conducting trainings similar to this one, included both a home-study component and an in-person training.

The home-study training component used an online distance learning platform. It provided trainees with an overview of the assessors' activities and covered topics such as working with school staff, parents, and students; conducting assessments via the computer; and carrying out timekeeping and expense report procedures. This component required assessors and team leaders

to log into the online training system, review the assigned modules, and complete and pass quizzes based on the material presented.

Subsequently, trainees attended a five-day, in-person training to receive step-by-step, comprehensive instruction on how to administer the assessments reliably. The training used a variety of teaching techniques, including lectures, round-robin demonstrations, and hands-on practice. Trainees participated in group and paired practice using scripted mock assessments. At the end of training, the assessors and field team leaders underwent a certification process, whereby a trainer observed and scored each trainee as he or she conducted an assessment with a child. Field team leaders received an additional day of training to review their management responsibilities.

During the data collection period, trainers conducted quality assurance visits in both the fall and spring to monitor the quality of the field staff's interactions with school staff and the technical aspects of administering the assessments. Each assessor and field team leader was observed once while assessing a student during each round of data collection. Any field staff member who was not following protocols received immediate feedback, additional training from a trainer or team leader, and additional quality assurance monitoring by the team leader.

Adaptive testing. Each of the assessments had features that made them adaptive—that is, the difficulty of the items was adjusted depending on the student's performance during the assessment. By concentrating testing time on items whose difficulty is appropriate to a student's ability, adaptive testing enhances test reliability per minute of testing time. It also decreases the likelihood of floor and ceiling effects, scenarios in which a student's ability is lower or higher than the range of abilities captured by an assessment. Given that our study sample was predominantly low-achieving (see Appendix A, Table A.6), adaptive testing was particularly important for ensuring that low achievers were administered assessment items appropriate to their skill level, enabling their skill level to be reliably measured.

When administering the language (CELF) and listening comprehension (WJ-III) assessments, assessors used approaches known as basal and ceiling rules to administer items of appropriate difficulty. Students were initially presented with items targeted to their age level and, if they did not answer a specified number of consecutive items correctly, assessors moved to progressively easier items until the student achieved the specified number of consecutive correct answers. At that difficulty level, known as the basal level, the student was assumed to be able to answer all easier items correctly and was therefore not administered those items. The assessor then administered increasingly more difficult items until the student gave a certain number of consecutive incorrect answers. At that difficulty level, known as the ceiling level, the student was assumed to be unable to answer any harder items, at which point the assessor stopped the test.

In both of the ECLS-K assessments (background knowledge and reading comprehension), the key adaptive feature was the use of two-stage assessments. The first stage of the assessment included items of a broad range of difficulty and was completed by all students. The second stage included multiple forms of the assessment with different levels of difficulty, and student's performance in the first stage determined which second-stage form was administered. The background knowledge assessment had two second-stage forms (low and high difficulty), and

the reading comprehension assessment had three second-stage forms (low, medium, and high difficulty).

Administration of the assessments also differed for students identified as speaking a non-English language at home. All students in the study were initially administered a language screener—consisting of two subtests (Simon Says and Art Show) from the Preschool Language Assessment Survey 2000 (Duncan and DeAvila 1998)—to assess their English proficiency. Students with a non-English home language who failed the language screener took only the assessment of basic language skills but no other assessment. This approach allowed for assessment of language growth for all students, including English language learners. At that same time, it reduced assessment burden on English language learners. Subjecting English language learners to the full battery of assessments could have led to scenarios in which they would not respond to enough of the items to establish a valid score, or would experience undue stress from being asked questions they did not understand. Students who passed the screener (regardless of home language) or who spoke English at home (regardless of screener performance) participated in all assessments appropriate for their grade level.

On each assessment, the potential items that a student could be administered were identical in the fall and spring. However, because each assessment had adaptive features, students would not necessarily have encountered the same items if their performance improved between the fall and spring. For example, if students demonstrated higher performance in the spring, then they could have been administered more difficult items before the assessment ended at their ceiling level (on the CELF and WJ-III) or even been administered a more difficult second-stage form of the assessment (on the ECLS-K assessments). Neither the students nor classroom teachers received copies of the assessment items and answers.

3. Calculating final scores from the assessments

In each assessment, a student's performance on the test items generated a final summary score that measured the student's ability in language or comprehension. The scores, known as theta scores, were obtained from item response theory, a method for placing all students on the same scale of ability even if they were not administered the same set of items. Theta scores from the fall and spring and across different grades were on the same scale (although this study did not compare students from different grades).

We obtained theta scores in one of several ways, depending on the assessment. The mathematical transformation of theta scores from the W-J III (listening comprehension) assessment, referred to as W scores by the test publisher, were derived from the number of items answered correctly between students' basal and ceiling levels, using software from the test publisher. For the two ECLS-K tests (background knowledge and reading comprehension), the Educational Testing Service, which originally developed the tests, generated theta scores from students' item responses using the same item response theory models specified in U.S. Department of Education (2002) and Pollack et al. (2005). For the CELF (basic language skills) assessment, theta scores were not directly available from the test publisher because we administered only a selection of the subtests from the assessment. Instead, the study team produced theta scores from the students' item responses on the CELF using an item response theory model known as a Rasch model (Rasch 1960), the same type of model used in the listening comprehension assessment.

In all analyses, we standardized the theta scores or W scores into *z*-scores by subtracting the fall study sample mean and dividing by the fall study sample standard deviation within each grade separately. Therefore, in both the fall and spring, each student's final *z*-score was expressed as the number of standard deviations above or below the average fall grade-level score in the study.

Earlier, in Appendix A, we reported the number of students in the final analysis sample—those who had a fall and spring score on at least one assessment (Appendix A, Table A.3). Among those students, Table B.7 shows the number of students who had fall and spring scores on each of the four domains tested by the study.

Table B.7. Number of students with fall and spring scores, by assessment

Group	Sample size
Prekindergarten and kindergarten	
Had both fall and spring score on	
Basic language skills	1,778
Background knowledge	1,697
Listening comprehension	1,716
Had both fall and spring score on at least one assessment (final analysis sample)	1,783
Grades 1 through 3	
Had both fall and spring score on	
Basic language skills	3,183
Background knowledge[a]	987
Listening comprehension	3,176
Reading comprehension[b]	2,094
Had both fall and spring score on at least one assessment (final analysis sample)	3,186

Source: Authors' calculations using data from the fall and spring tests administered by the study team.

[a] Grade 1 was the highest grade that was administered the background knowledge assessment.

[b] The reading comprehension assessment was administered in grades 2 and 3 only.

4. Reliability and validity of the assessments in national student samples

The developers of the assessments in this study previously administered those assessments to nationally representative samples of students to estimate national distributions of achievement (known as norms) for students of each age or grade. Based on the scores from those norming samples, the test developers evaluated several psychometric properties of the assessments, including their reliability and validity.

All assessments had evidence of reliability and validity (Table B.8). For example, most of the assessments—and, for the CELF, most of the subtests—had internal consistency (alpha) values of at least 0.7 or higher. Test publishers typically demonstrated the validity of the assessments by showing that scores on those assessments were at least moderately (0.4 or above) or highly (0.7 or above) correlated with scores on other previously validated assessments measuring similar types of skills.

Table B.8. Reliability and validity of the assessments in national student samples

Assessment	Evidence of reliability and validity
	Basic language skills
CELF P–2 (Wiig et al. 2004)	**Reliability:** ***Internal consistency (alpha)*:** Varies by subtest and age (ages 3–6), ranging from 0.78 to 0.85 for Concepts and Following Directions, 0.70 to 0.84 for Expressive Vocabulary, 0.90 to 0.95 for Word Classes, and 0.78 to 0.83 for Sentence Structure except for 0.69 among 6-year-olds. ***Test-retest reliability*:** Ranges from 0.75 to 0.95 across subtests for ages 3 to 5; 0.63 to 0.69 across subtests for age 6 (three of four subtests). ***Interrater reliability*:** 0.95 for the expressive part of Word Classes. **Validity:** Moderate to high correlations with the first edition of the assessment (0.50 to 0.68 for subtest scores), CELF–4 (0.59 to 0.85 for subtest scores), and the Preschool Language Scale–Fourth Edition (0.71 to 0.72 for composite scores).
CELF–4 (Semel et al. 2003)	**Reliability:** ***Internal consistency (alpha)*:** Varies by subtest and age (ages 5–9), ranging from 0.81 to 0.92 for Concepts and Following Directions, 0.80 to 0.85 for Expressive Vocabulary, 0.74 to 0.91 for Word Classes, and 0.64 to 0.76 for Sentence Structure. ***Test-retest reliability*:** Ranges from 0.69 to 0.91 for subtest scores, except for 0.49 among 7-year-olds' Sentence Structure scores. ***Interrater reliability*:** 0.95 for the expressive part of Word Classes. **Validity:** Moderate to high correlations with the third edition of the assessment (0.81 for Concepts and Following Directions, 0.68 for Word Classes, and 0.55 for Sentence Structure). There are no correlation scores to the CELF-3 for Expressive Vocabulary because it was a new subtest added to the CELF-4.
	Background knowledge
ECLS–K General Knowledge (U.S. Department of Education 2002)	**Reliability:** ***Internal consistency (alpha):*** Ranges from 0.78 to 0.79 for the first-stage routing form and 0.64 to 0.74 for the second-stage skill level forms. ***Reliability of the theta score*:** Ranges from 0.88 to 0.89 across time points. **Validity:** Moderate correlations (0.57 to 0.59) with the ECLS–K reading score at grades 1, 3, and 5 (Claessens et al. 2009; Duncan et al. 2007).
	Listening comprehension
W-J III Oral Comprehension subtest (Woodcock et al. 2001, 2007)	**Reliability:** ***Internal consistency*** (split-half reliability for ages 4 through 7): Ranges from 0.78 to 0.88 (McGrew et al. 2007). **Validity:** Moderate correlations (0.45 to 0.59) with reading and language subtest scores of the Kaufman Test of Education Achievement and the Wechsler Individual Achievement Test (McGrew and Woodcock 2001).
	Reading comprehension
ECLS–K Third Grade Reading Assessment (U.S. Department of Education 2004; Pollack et al. 2005)	**Reliability:** ***Internal consistency (alpha)*:** 0.75 for first-stage routing form and 0.79 to 0.84 for second-stage skill level forms. ***Reliability of the theta score*:** 0.94. **Validity:** High correlation (0.83) with the Woodcock-McGrew-Werder Mini-Battery of Achievement total score across reading and math subtests in grades 2 and 3.

Source: Publications cited in the table.

CELF P–2 = Clinical Evaluation of Language Fundamentals Preschool–Second Edition; CELF–4 = Clinical Evaluation of Language Fundamentals–Fourth Edition; ECLS–K = Early Childhood Longitudinal Study–Kindergarten Class of 1998–99; preLAS = Preschool Language Assessment Survey 2000; W-J III = Woodcock-Johnson III, Tests of Achievement.

5. Variation and reliability of test scores in the study sample

The reliability and validity evidence in Table B.8, based on national student samples examined by the test publishers, was not guaranteed to apply to this study sample. As discussed in Appendix A, students in the study demonstrated considerably lower performance than the average student nationwide. For a majority of assessments and grade spans, the median student in the study scored at no more than the 30th percentile in the national population (Appendix A, Table A.6).

It was therefore important to determine whether these assessments could make reliable distinctions among students with different skill levels even within this generally low-achieving study sample. To do so, we documented two key characteristics of the test scores in the study: (1) the extent to which they differed across students and (2) the extent to which these differences were reliable—that is, the degree to which they would be consistently observed in repeated measurements based on this assessment, rather than reflecting transitory measurement error.

Test scores in the study differed substantially across students. Appendix A, Table A.6 documented the degree of variation in the scores. Across assessments, grade spans, and points in time, high performers in the study—those who outperformed 80 percent of other study students in the same age or grade—typically scored at the 50th percentile or above within the national population. In contrast, low performers in the study—those who performed worse than 80 percent of other study students in the same grade—scored at less than the 20th percentile within the national population.

These differences in test scores were also highly reliable. For the three assessments in which item response theory models were directly estimated on the study's item-level data—the CELF (basic language skills), ECLS-K General Knowledge (background knowledge), and ECLS-K Third Grade Reading (reading comprehension) assessments—we could calculate the reliability of the theta scores. Those reliability values ranged from 0.89 to 0.97 (Table B.9).

Table B.9. Reliability of test scores in the study

Domain and assessment	Reliability of the theta score
Basic language skills (CELF P–2 and CELF–4)	
Fall 2011 and spring 2012, pooled	0.97
Background knowledge (ECLS-K General Knowledge)	
Fall 2011	0.89
Spring 2012	0.91
Reading comprehension (ECLS-K Third Grade Reading)	
Fall 2011	0.93
Spring 2012	0.94

Source: Authors' calculations using data from the fall and spring tests administered by the study team.

C. Measuring background characteristics of teachers and students

Two surveys of teachers provided information about the background characteristics of the study participants.[11] First, a ***teacher self-report*** asked teachers to report information about their own background characteristics and classroom experiences. Second, a ***teacher student report*** asked teachers to report the characteristics of their students who were in the study. Both surveys were conducted in the spring of 2012. Teachers were offered a $20 gift card incentive upon completion of the self-report and an additional $5 added to that gift card for each student report completed. Earlier, in Appendix A, we reported the number of classrooms and students in the final analysis sample (Appendix A, Table A.3). Within this final analysis sample, Tables B.10 and B.11 show the number of classrooms with completed teacher self-reports and the number of students with completed teacher student reports. The remainder of this section provides a brief overview of the content and modes of data collection for these surveys.

Table B.10. Number of classrooms with completed teacher self-reports

Group	Number of classrooms
Prekindergarten and kindergarten	
In final analysis sample	378
Completed teacher self-report	339
Grades 1 through 3	
In final analysis sample	657
Completed teacher self-report	577

Source: Authors' calculations from study-collected sample information and teacher self-reports.

Table B.11. Number of students with completed teacher student reports

Group	Number of students
Prekindergarten and kindergarten	
In final analysis sample	1,783
Had a completed teacher student report	1,550
Grades 1 through 3	
In final analysis sample	3,186
Had a completed teacher student report	2,680

Source: Authors' calculations from study-collected sample information and teacher student reports.

1. Teacher self-report

We asked the teachers in the study to complete a 30-minute, web-based self-report in the spring of 2012. Questions that asked teachers to report their professional and demographic characteristics provided contextual information for identifying the types of teachers to which the study findings might be most relevant (see Appendix A, Table A.5). Other topics on the survey, which were not the focus of this report, included the teacher's use of curricula, involvement in instructional leadership activities, strategies for instructional planning, and approaches to reading and language arts instruction.

[11] We also surveyed principals and prekindergarten directors. Data from and information about those surveys can be found in the study's restricted-use file and the accompanying documentation.

2. Teacher student report

We asked teachers to complete a web-based survey about each of the study students they taught. This instrument, the teacher student report, was the study's source of information about students' participation in support services, such as English as a Second Language and special education. We controlled for these student background characteristics when estimating teachers' contributions to student growth (see Appendix C). Other topics on the teacher student report, which were not the focus of this report, included the students' absences, socio-emotional traits, and academic skills, and the level of involvement of the students' parents or guardians during the school year.

This page has been left blank for double-sided copying.

APPENDIX C: ANALYTIC METHODS

The objective of the study's analysis was to identify instructional practices associated with teachers' contributions to student growth in language and comprehension. The key steps in the analysis involved (1) creating summary measures of instructional practices, (2) measuring teachers' contributions to student growth, and (3) assessing the relationships between the summary measures of practices and teachers' contributions to student growth. This appendix describes the technical details for each of these steps.

A. Creating summary measures of instructional practices

The 285 items on the study's observation instrument, the Observation of Language and Literacy Instruction (OLLI), captured many specific aspects of instruction. Examining the relationship between each of these items and student growth would have led to many imprecisely estimated relationships. This would make it difficult to extract clear hypotheses on the most promising ways to promote language and comprehension growth. To sharpen the study's focus on a smaller number of instructional practices, we used data-driven approaches to identify groups of items that were strongly related to each other because they reflected the same underlying instructional practice. Each group of items formed a summary measure of an instructional practice that could be examined in subsequent analyses of relationships with student growth. Following the steps described in this section (and summarized in the main report in Figure II.1), we began with data on 285 items measured for each observation session (averaged across the six segments that made up each session). This process resulted in 13 summary measures of instructional practices measured at the classroom level.

1. Adjust item scores for differences among observers

Systematic differences in how observers scored items, referred to as observer effects, could generate differences in item scores across classrooms that did not reflect true differences in practices. To address this potential problem, we assessed whether observers systematically differed in the scores they assigned, and then adjusted item scores to remove those observer effects.

Assess the presence of observer effects. To determine whether observers differed in the way they scored items, we assessed whether, on each item, any portion of the variation in scores was due to differences across observers. To do this, we leveraged the fact that each observer conducted observations in multiple classrooms, and each classroom was observed by multiple observers. Conceptually, if an observer assigned higher scores to a classroom than did other observers who rated the same classroom, the higher scores might simply reflect chance factors that led the teacher to perform unusually well in that observation session. However, if this observer consistently assigned unusually high scores in multiple classrooms, then this scenario would provide evidence that the observer was systematically more lenient than others. To assess the presence of observer effects, we estimated cross-classified random effects models—one for each OLLI item—that decomposed the total variation in scores across all observation sessions into portions due to differences across schools, observers, classrooms within schools, and sessions within classrooms (Luo and Kwok 2009; Meyers and Beretvas 2006).

On average, 14 percent of the variation in item scores was due to observer effects. Therefore, this decomposition indicated that, on a typical item, some observers gave systematically higher scores than did others. The largest source of variation occurred across observation sessions within classrooms (79 percent). Other sources of variation consisted of differences across schools (2 percent) and classrooms within schools (6 percent).

Adjust for observer effects. Because we found differences in item scoring across observers, we adjusted item scores to remove those differences using a regression-based approach (Raymond and Viswesvaran 1993; Houston et al. 1991). We used session-level data to estimate an ordinary least squares regression separately for each OLLI item, with item scores as the dependent variable and a full set of observer indicators (binary variables, one for each observer) and classroom indicators (binary variables, one for each classroom) as the independent variables. We controlled for the classroom indicators to account for the possibility that some observers may have observed more effective teachers than other observers, and such true differences in instructional quality should not be interpreted as observer effects. After estimating each regression, we subtracted each observer's unique effect from the item scores of the sessions that the observer observed. After this adjustment, none of the remaining variation in item scores across observation sessions could be attributable to systematic differences in item scoring among observers.

This adjustment removed the influence of biases or errors that observers consistently demonstrated across all of their sessions. However, it did not remove the influence of observers' biases or errors that were specific to particular sessions—known as observer-by-session interactions. Those interactions remained a source of measurement error that we addressed using steps described later (see Section A, subsection 4 of this appendix for a description of how we used empirical Bayes methods to prevent measurement error from biasing the estimated relationships between practices and student growth).

2. Create composite items

As discussed earlier, our main objective in the analysis of the OLLI data was to identify a smaller number of well-defined instructional practices that underlay the large number of OLLI items. However, standard techniques to identify underlying behaviors from observed items, such as exploratory factor analysis, could not have incorporated such a large number of OLLI items—285 in total. With such a large number of items, the behaviors identified by a factor analysis would be expected to fit the data poorly (Marsh et al. 2014). For this reason, before attempting to identify well-defined instructional practices, we first reduced the number of items in two ways: (1) removing non-instructional items and (2) combining some closely-related items into composite items.

Exclude non-instructional items. We removed items that were not intended to capture instruction. Specifically, we removed items meant only to describe the classroom context (such as the number of adults and children in the room and how children were being grouped) rather than a specific aspect of instruction. We also removed items that were redundant with other included items. The redundant items were those that measured the absence of a given behavior (such as the absence of approaches to engaging students) while the included items were those that measured the presence of various types of that behavior (such as different approaches to

engaging students). The absence of a behavior could be directly inferred in cases when none of the different types of that behavior was present.

Create composite items. Some items were well-suited for being combined into composite items because they pertained to the same category of teacher or student actions and were listed together under the same or similar prompt in the OLLI. For example, one list consisted of a series of items in the OLLI prompted by the question, "What techniques did the teacher use to help students expand their use of language?" with each item pertaining to a single technique. The OLLI contained four such lists, discussed in more detail below.

Within each list of items, we created composite items empirically using principal components analysis (PCA). PCA is an empirical technique for creating composite items, known as principal components, as a linear combination of the original items, such that the composite items retain as much of the variance in scores from the original items as possible. Because this stage of the analysis did not yet entail identifying well-defined instructional practices, it was important to retain as much of the original variation in item scores as possible so that this information could be used in subsequent stages of analysis to identify instructional practices. For this reason, PCA was particularly suited to creating composite items. All item scores that we submitted to the PCA were still measured at the level of the observation session (averaged across segments within each session), so the resulting composite items were also measured at the session level.

When deciding how many principal components to extract from a list of items, we took into account three key considerations. First, we examined how much total variance in the original items was explained by each principal component, called the eigenvalue. Eigenvalues decreased with each additional principal component. In an approach called a scree test, we generally stopped extracting principal components right before the last substantial drop in eigenvalues (Cattell 1966). Second, we looked at the standardized component loadings—the predicted change in an item score (in standard deviation units) associated with a one standard deviation change in a principal component. We required that every principal component we extracted should have at least three items with salient loadings, defined as loadings of at least 0.30. This requirement helped ensure that the principal components could explain a meaningful amount of item variance. Third, we gave priority to principal component solutions that had simple structures in which each item loaded saliently onto only one component, which helped enhance the interpretability of the principal components.

We did not consider the internal consistency (typically measured by Cronbach's alpha) of the principal components when deciding how many principal components to extract. Internal consistency of principal components would have been important if they had been used as the final summary measures of instructional practices. In this study, however, the principal components were considered only as individual items that would potentially contribute (along with other items from the OLLI) to the final summary measures. Only the final summary measures needed to have sufficient internal consistency. In addition, we did not require that the principal components collectively account for a minimum amount of total item variance; given the study's exploratory purpose, it was acceptable for some items not to contribute strongly to any principal components.

For any list of items from which we extracted two or more principal components, there was an infinite number of alternative solutions that could explain the item variance equally well (Fabrigar et al. 1999). To choose one of these solutions, we followed the widely accepted practice of selecting one with a simple structure—one with each item tending to load saliently onto only one principal component—to make the principal components easily interpretable (Thurstone 1947). The technique we used to identify a simple structure, promax rotation (Hendrickson and White 1964), did not require the final principal components to be uncorrelated, which had the advantage of minimizing the assumptions imposed on the data.

As noted earlier, we applied PCA separately to each of four lists of items that shared the same, or a similar, prompt. We created a total of 12 composite items from these four lists, as follows:

- **Items on expanding students' use of language.** Six items shared the prompt, "What techniques did the teacher use to help students expand their use of language?" From these items, the PCA generated one composite item, which measured the frequency and diversity of techniques to help expand students' use of language.
- **Items on engaging students.** Seventeen items shared the prompt, "In what ways did the teacher engage students in activities?" or "In what ways did the teacher encourage student interaction?" From these items, the PCA generated four composite items that measured (1) engaging students through games and hands-on activities, (2) encouraging students to speak and read with peers, (3) engaging students through writing activities, and (4) engaging students through teacher-directed or choral activities.
- **Items on what the teacher talked about in reading activities.** Thirty-six items shared the prompt, "What did the teacher talk/ask about during [pre-reading, reading, or post-reading]?" From these items, the PCA generated six composite items that measured (1) the frequency and diversity of pre-reading activities, (2) focusing on meaning, vocabulary, and comprehension strategies during reading, (3) the frequency and diversity of post-reading activities, (4) teaching letters, words, grammar, and spelling, (5) teaching text features, and (6) focusing on the purpose of a text and activating prior knowledge.
- **Items on approaches for teaching world knowledge.** Eight items shared the prompt, "What approaches did the teacher use to introduce, reinforce, or teach world knowledge?" From these items, the PCA generated one composite item that measured the frequency and diversity of world knowledge activities.

Tables C.1 through C.12 specify the OLLI items that loaded saliently onto each of the 12 principal components and provide descriptive statistics for those items. Within each table, OLLI items are ordered from highest to lowest component loading. The few items that loaded saliently onto more than one principal component are listed in more than one table. Seven items that did not load saliently onto any principal component are not listed in any table. In all of these tables, average scores for each item are calculated across observation sessions. For example, on average across observation sessions, the teacher talked about the title, topic, subject, or theme of the text during post-reading in 2 percent of observation segments (Table C.8). (As discussed in Appendix B, each observation session consisted of six 15-minute segments, for a total of 90 minutes of observed instruction per session.)

Within each list of items described earlier, all of the items in the list had some degree of association—even if only a very small association—with all of the principal components generated from that list. In formal terms, all items in a list had nonzero loadings onto all principal components generated from that list. For example, in the list of items on what the teacher talked about in reading activities, all 36 items had nonzero loadings onto all six principal components that the PCA generated from that list. However, when a loading was not salient (that is, did not reach 0.30), a principal component could explain very little of the variation in the item (in fact, less than 9 percent of the variation if all principal components were hypothetically uncorrelated). For this reason, Tables C.1 through C.12 do not show items with nonsalient loadings onto the principal components.

To calculate each observation session's score on a composite item, we obtained the component score on the principal component represented by that composite. The component score was calculated with the regression method (Thurstone 1935) using all of the individual items that were submitted to the PCA and their exact component loadings, even if some of those loadings were not salient.

Table C.1. Composite item on the frequency and diversity of techniques to help expand students' use of language: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
Teacher allowed students time to respond to questions	0	1	0.76	0.87
Teacher asked open-ended questions or questions that help students say more	0	1	0.68	0.87
Teacher added more information to what the student said	0	1	0.58	0.77

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

Table C.2. Composite item on engaging students through games and hands-on activities: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
Teacher had students work with peers on a hands-on activity	0	1	0.07	0.82
Teacher had students engage in a hands-on activity	0	1	0.13	0.81
Teacher had students play games	0	1	0.06	0.53
Teacher allowed informal student interactions/talk	0	1	0.45	0.48

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

Table C.3. Composite item on encouraging students to speak and read with peers: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
Teacher had students speak with each other	0	1	0.14	0.84
Teacher had students briefly discuss with peers (four minutes or less)	0	1	0.08	0.69
Teacher had students discuss with peers for more than four minutes	0	1	0.04	0.55
Teacher had students read with partners	0	1	0.03	0.34

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

Table C.4. Composite item on engaging students through writing activities: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
Teacher had students write a sentence or more	0	1	0.15	0.72
Teacher had students write about the topic, characters, or ideas in a book/text	0	1	0.05	0.68
Teacher had students use a book/text as a model for their writing	0	1	0.03	0.67

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

Table C.5. Composite item on engaging students through teacher-directed or choral activities: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
Teacher asked students questions	0	1	0.80	0.77
Students listened to the teacher or read silently	0	1	0.79	0.66
Teacher had students draw, act, or sing, or invited them to read along	0	1	0.46	0.60

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

Table C.6. Composite item on the frequency and diversity of pre-reading activities: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
During pre-reading, teacher talked about what the text may be about	0	1	0.04	0.69
During pre-reading, teacher talked about reading comprehension strategies	0	1	0.04	0.66
During pre-reading, teacher talked about the characters in the text	0	1	0.03	0.63
During pre-reading, teacher talked about the title, topic, subject, or theme of the text	0	1	0.10	0.62
During pre-reading, teacher connected the content with students' prior knowledge and experiences[d]	0	1	0.05	0.60
During pre-reading, teacher talked about key features of the book/text (type of book, parts of the book, author)	0	1	0.05	0.57
During pre-reading, teacher talked about vocabulary	0	1	0.05	0.53
During pre-reading, teacher talked about the text structure (parts of the story/text)[c]	0	1	0.01	0.37
During pre-reading, teacher talked about the purpose for reading the text[d]	0	1	0.04	0.36
Teacher announced the beginning of the reading activity[a]	0	1	0.19	0.35
During pre-reading, teacher talked about letters or words (sounding out letters or words, rhyming words, word recognition)[b]	0	1	0.03	0.32

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

[a] This item also contributed to the composite item on focusing on meaning, vocabulary, and comprehension strategies during reading (Table C.7).

[b] This item also contributed to the composite item on teaching letters, words, grammar, and spelling (Table C.9).

[c] This item also contributed to the composite item on teaching text features (Table C.10).

[d] This item also contributed to the composite item on focusing on the purpose of a text and activating prior knowledge (Table C.11).

Table C.7. Composite item on focusing on meaning, vocabulary, and comprehension strategies during reading: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
During reading, teacher talked about what happened in the story or what might happen next, or what information was presented in the text	0	1	0.12	0.75
During reading, teacher talked about the characters in the text	0	1	0.08	0.66
During reading, teacher connected content with students' prior knowledge and experiences[c]	0	1	0.07	0.57
During reading, teacher talked about the title, topic, subject, or theme of the text	0	1	0.06	0.54
During reading, teacher talked about vocabulary	0	1	0.10	0.53
During reading, teacher talked about reading comprehension strategies	0	1	0.06	0.48
Teacher announced the end of the reading activity	0	1	0.10	0.43
During reading, teacher talked about letters or words (sounding out letters or words, rhyming words, word recognition)[b]	0	1	0.09	0.38
Teacher announced the beginning of the reading activity[a]	0	1	0.19	0.32
During reading, teacher engaged in talk that was related to the text but not about its topic or content	0	1	0.06	0.31

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

[a] This item also contributed to the composite item on the frequency and diversity of pre-reading activities (Table C.6).

[b] This item also contributed to the composite item on teaching letters, words, grammar, and spelling (Table C.9).

[c] This item also contributed to the composite item on focusing on the purpose of a text and activating prior knowledge (Table C.11).

Table C.8. Composite item on the frequency and diversity of post-reading activities: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
During post-reading, teacher talked about what the text was about	0	1	0.06	0.68
During post-reading, teacher talked about the characters in the text	0	1	0.04	0.67
During post-reading, teacher talked about reading comprehension strategies	0	1	0.03	0.62
During post-reading, teacher talked about the title, topic, subject, or theme of the text	0	1	0.02	0.60
During post-reading, teacher talked about the text structure (parts of the story/text)[b]	0	1	0.01	0.50
During post-reading, teacher talked about vocabulary[a]	0	1	0.03	0.50
During post-reading, teacher talked about evaluating the text	0	1	0.02	0.47
During post-reading, teacher talked about key features of the text (type of book, parts of the book, author)[b]	0	1	0.01	0.44
During post-reading, teacher talked about the purpose for reading the text[c]	0	1	0.02	0.42

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

[a] This item also contributed to the composite item on teaching letters, words, grammar, and spelling (Table C.9).

[b] This item also contributed to the composite item on teaching text features (Table C.10).

[c] This item also contributed to the composite item on focusing on the purpose of a text and activating prior knowledge (Table C.11).

Table C.9. Composite item on teaching letters, words, grammar, and spelling: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
During pre-reading, teacher talked about grammar, mechanics, or spelling	0	1	0.02	0.65
During reading, teacher talked about grammar, mechanics, or spelling	0	1	0.05	0.65
During post-reading, teacher talked about grammar, mechanics, or spelling	0	1	0.02	0.62
During reading, teacher talked about letters or words (sounding out letters or words, rhyming words, word recognition)[b]	0	1	0.09	0.60
During post-reading, teacher talked about letters or words (sounding out letters or words, rhyming words, word recognition)	0	1	0.02	0.55
During pre-reading, teacher talked about letters or words (sounding out letters or words, rhyming words, word recognition)[a]	0	1	0.03	0.54
During post-reading, teacher talked about vocabulary[c]	0	1	0.03	0.30

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

[a] This item also contributed to the composite item on the frequency and diversity of pre-reading activities (Table C.6).

[b] This item also contributed to the composite item on focusing on meaning, vocabulary, and comprehension strategies during reading (Table C.7).

[c] This item also contributed to the composite item on the frequency and diversity of post-reading activities (Table C.8).

Table C.10. Composite item on teaching text features: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
During reading, teacher talked about the text structure (parts of the story/text)	0	1	0.02	0.70
During pre-reading, teacher talked about the text structure (parts of the story/text)[a]	0	1	0.01	0.55
During reading, teacher talked about key features of the text (type of book, parts of the book, author)	0	1	0.02	0.55
During post-reading, teacher talked about the text structure (parts of the story/text)[b]	0	1	0.01	0.46
During post-reading, teacher talked about key features of the text (type of book, parts of the book, author)[b]	0	1	0.01	0.35

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

[a] This item also contributed to the composite item on the frequency and diversity of pre-reading activities (Table C.6).

[b] This item also contributed to the composite item on the frequency and diversity of post-reading activities (Table C.8).

Table C.11. Composite item on focusing on the purpose of a text and activating prior knowledge: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
During reading, teacher talked about the purpose for reading the text	0	1	0.03	0.65
During pre-reading, teacher talked about the purpose for reading the text[a]	0	1	0.04	0.58
During post-reading, teacher talked about the purpose for reading the text[c]	0	1	0.02	0.49
During reading, teacher connected content with students' prior knowledge and experiences[b]	0	1	0.07	0.34
During pre-reading, teacher connected content with students' prior knowledge and experiences[a]	0	1	0.05	0.30

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

[a] This item also contributed to the composite item on the frequency and diversity of pre-reading activities (Table C.6).

[b] This item also contributed to the composite item on focusing on meaning, vocabulary, and comprehension strategies during reading (Table C.7).

[c] This item also contributed to the composite item on the frequency and diversity of post-reading activities (Table C.8).

Table C.12. Composite item on the frequency and diversity of world knowledge activities: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Component loading
	Minimum	Maximum		
Teacher and/or students reviewed or discussed facts about world knowledge	0	1	0.29	0.77
Teacher presented detailed information about a world knowledge topic	0	1	0.10	0.72
Teacher and/or students provided a definition of a word or concept related to world knowledge	0	1	0.13	0.70
Teacher and/or student named or listed things (objects, places, events, actions, people)	0	1	0.21	0.57
Teacher read to students about a world knowledge topic	0	1	0.06	0.54
Teacher had students read about a world knowledge topic	0	1	0.05	0.47
Teacher and/or students used technology or multimedia in a world knowledge activity	0	1	0.06	0.33

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. An item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest component loading. Only items with component loadings of at least 0.30 are listed (even though other items may have loaded onto the principal component with lower loadings).

This process resulted in a reduced set of 89 items—12 composite items, plus 77 original items that were not incorporated into composites. On the one hand, 89 items, if analyzed individually for relationships with student growth, would still yield a large number of imprecisely estimated relationships with few clear lessons. On the other hand, these items were sufficiently reduced in number to permit standard techniques to identify coherent groups of items representing the same underlying instructional practice. We describe next the process for identifying the underlying practices that served as the focus of the remainder of the study.

3. Construct summary measures of practices for each observation session

Analytic procedure. To measure a smaller number of instructional practices, we used exploratory factor analysis (EFA) to identify groups of items that were highly correlated with each other because they reflected a common "factor"—a well-defined instructional practice. Each group of items gave rise to a summary measure of an instructional practice.

We chose EFA rather than PCA to create the final summary measures because EFA focused only on the variation in scores that items shared with each other. Therefore, EFA was well-suited to identifying instructional practices that could explain why multiple items were related to each other. The aim was not necessarily to retain the greatest possible variation in scores from the original items (including variation that was not shared with other items)—a task that would have been suited to PCA. Moreover, we chose not to impose restrictions on the EFA based on any

conceptual framework—for instance, by using theory to specify which items have the potential to contribute to the same summary measure. Adopting a purely empirical approach to identifying the underlying instructional practices was consistent with the exploratory nature of this study—letting the data reveal, to the maximum extent possible, the practices that teachers were using.

From the 89 available items (77 individual items and the 12 composite items described above), we first assessed the likely number of factors underlying those items. To do so, we used a technique called minimum average partialling, which successively identified factors and removed the variance of item scores associated with those factors until the average correlation between all items was minimized (Velicer 1976; Velicer et al. 2000). This technique suggested that the items reflected anywhere from 13 factors (based on the method in Velicer [1976]) to 17 factors (based on the method in Velicer et al. [2000]).

To consider a comprehensive set of possible factor solutions, we generated factor solutions for each scenario in which the number of factors ranged from 1 to 15. (We did not generate factor solutions with more than 15 factors because the 14-factor and 15-factor solutions already contained some factors with which no items were strongly related according to the criteria described below.) In each scenario, we used data at the level of the observation session to estimate the EFA model using principal axis factoring, a procedure that avoided the assumption of multivariate normality in the item scores (Fabrigar et al. 1999). As in the PCA described earlier, the EFA used oblique (promax) rotation to enhance the likelihood of obtaining a factor solution with a simple structure while not requiring the factors to be uncorrelated. All estimates used analysis weights that took into account the study's sampling design and pattern of nonresponse (see Appendix A). We dropped a very small number of observation sessions—19 of 4,113 sessions, or 0.5 percent—due to missing data on at least one of the OLLI items.

Among these possible factor solutions, we sought to choose the best solution based on four criteria. First, the factors needed to have an acceptable level of internal consistency, with Cronbach's alpha exceeding 0.70. Second, we considered the items' standardized factor loadings—the predicted change in an item score (in standard deviation units) for a one standard deviation change in a factor. Each factor needed to have at least three items with salient loadings, defined as a loading of at least 0.30. Third, no items could load saliently onto more than one factor, allowing the factor solution to exhibit a simple structure. Fourth, each factor needed to have a well-defined interpretation.

Among the 15 possible solutions, we chose the 13-factor solution because it satisfied all of the criteria described above. These 13 factors represented the 13 instructional practices examined in the main report. Chapter II of the main report (Table II.2) listed and briefly described these instructional practices. Here, in Tables C.13 through C.25, we specify the OLLI items that loaded saliently onto each instructional practice and provide descriptive statistics for those items. Within each table, OLLI items are ordered from highest to lowest factor loading. Eleven items that did not load saliently onto any instructional practice are not listed.

All 89 items had some degree of association—even if only a very small association—with all 13 instructional practices. However, when a factor loading was not salient (that is, did not reach 0.30), an instructional practice could explain very little of the variation in the item (in fact,

less than 9 percent of the variation if all instructional practices hypothetically were uncorrelated). For this reason, Tables C.13 through C.25 do not show items with nonsalient loadings.

Using the factor structure shown in Tables C.13 through C.25, each observation session was assigned a score (called a factor score) on each of the 13 instructional practices. These factor scores were the initial summary measures of the practices observed in each observation session. We obtained factor scores by applying the regression method (Thurstone 1935) using all of the items that were submitted to the EFA and their exact factor loadings, even if some of those loadings were not salient.

Although the EFA (and PCA in the previous step) did not take into account the clustering of the data—with observation sessions clustered within classrooms, and classrooms clustered within schools—these analyses still produced valid (consistent) estimates of the factor or component loadings (Muthén 1991). Failure to account for clustering would have led to erroneous tests of the statistical significance of the loadings and erroneous tests of model fit. However, the procedures described in this section used only the estimated loadings without employing any tests of statistical significance. For this reason, there was no need to account for clustering at these stages of the analysis.

Table C.13. Encouraging students' oral language: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Teacher's talk was mostly for instruction or content	0	1	0.71	0.69
Number of minutes (out of 15) in which teacher was talking with students (average across segments)	2.5	15	13.68	0.68
Composite item: frequency and diversity of techniques to help expand students' use of language	NL	NL	0.04	0.62
Composite item: engaging students through teacher-directed or choral activities	NL	NL	0.03	0.61
Level of teacher's enthusiasm on 0-to-2 scale (average across segments)	0	2	1.68	0.55
Teacher's frequency of interaction with students on 1-to-3 scale (average across segments)	1	3	2.68	0.53
Fraction of students on whom the teacher called (average across segments)	0	1	0.66	0.51
Fraction of students who spoke with teacher (average across segments)	0	1	0.50	0.47
Teacher's talk was mostly giving directions	0	1	0.19	0.41
Teacher's talk was mostly on behavior management	0	1	0.05	0.40
Clarity and distinctness of the teacher's speech on 0-to-2 scale (average across segments)	0	2	1.95	0.31

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

NL is no limit.

Table C.14. Focusing on phonics and grammar during reading: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Teacher and students did not define words before reading	0	1	0.19	0.79
Teacher and students did not define words during reading	0	1	0.16	0.77
Teacher and students did not define words after reading	0	1	0.14	0.70
Composite item: teaching letters, words, grammar, and spelling	NL	NL	-0.06	0.36
Number of texts taught (average across segments)	0	NL	0.76	0.32
When reading out loud, teacher emphasized things other than the content or subject of the text (such as word sounds or sentence structure)	0	1	0.12	0.30

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

NL is no limit.

Table C.15. Engaging students in defining new words during pre-reading: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Teacher or students used more than one approach to define a word during pre-reading	0	1	0.03	0.88
Extent of students' involvement in defining words during pre-reading on 0-to-3 scale (average across segments)	0	3	0.09	0.84
Teacher or students defined words by providing additional descriptors during pre-reading	0	1	0.03	0.80
Teacher or students provided a definition of a word during pre-reading	0	1	0.06	0.76
Teacher or students defined words by showing a picture or using a gesture or vocal quality during pre-reading	0	1	0.02	0.63

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

Table C.16. Engaging students in defining new words during reading: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Teacher or students used more than one approach to define a single word during reading	0	1	0.04	0.88
Extent of students' involvement in defining words during reading on 0-to-3 scale (average across segments)	0	3	0.11	0.83
Teacher or students defined words by providing additional descriptors during reading	0	1	0.04	0.76
Teacher or students provided a definition of a word during reading	0	1	0.07	0.76
Teacher or students defined words by showing a picture or using a gesture or vocal quality during reading	0	1	0.03	0.61

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

Table C.17. Engaging students in defining new words during post-reading: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Teacher or students used more than one approach to define a single word during post-reading	0	1	0.01	0.89
Teacher or students defined words by providing additional descriptors during post-reading	0	1	0.01	0.83
Extent of students' involvement in defining words during post-reading on 0-to-3 scale (average across segments)	0	3	0.04	0.82
Teacher or students provided a definition of a word during post-reading	0	1	0.02	0.76
Teacher or students defined words by showing a picture or using a gesture or vocal quality during post-reading	0	1	0.01	0.62

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

Table C.18. Engaging students in defining new words outside of reading: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Teacher or students defined words by providing additional descriptors outside of reading	0	1	0.08	0.79
Teacher or students provided a definition of a word outside of reading	0	1	0.16	0.78
Students had some involvement in defining words outside of reading	0	1	0.07	0.66
Teacher or students defined words by showing a picture or using a gesture or vocal quality outside of reading	0	1	0.07	0.61
Students had extended involvement in defining words outside of reading	0	1	0.02	0.45
Students had minimal involvement in defining words outside of reading	0	1	0.06	0.43
Students listened to teacher define words outside of reading	0	1	0.08	0.41

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

Table C.19. Focusing on the meaning of texts during pre-reading: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Extent to which teacher organized talk about the content of a text during pre-reading on 0-to-2 scale (average across segments)	0	2	0.13	0.92
Extent of detail that teacher used to talk about the content of a text during pre-reading on 0-to-2 scale (average across segments)	0	2	0.12	0.92
Composite item: frequency and diversity of pre-reading activities	NL	NL	0.00	0.67

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

NL is no limit.

Table C.20. Focusing on the meaning of texts during reading: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Extent of detail that teacher used to talk about the content of a text during reading on 0-to-2 scale (average across segments)	0	2	0.26	0.93
Extent to which teacher organized talk about the content of a text during reading on 0-to-2 scale (average across segments)	0	2	0.26	0.90
Composite item: focusing on meaning, vocabulary, and comprehension strategies during reading	NL	NL	0.03	0.72
When reading out loud, teacher emphasized things related to the content or subject of the text	0	1	0.18	0.40

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

NL is no limit.

Table C.21. Focusing on the meaning of texts during post-reading: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Extent of detail that teacher used to talk about the content of a text during post-reading on 0-to-2 scale (average across segments)	0	2	0.12	0.94
Extent to which teacher organized talk about the content of a text during post-reading on 0-to-2 scale (average across segments)	0	2	0.12	0.92
Composite item: frequency and diversity of post-reading activities	NL	NL	0.02	0.69

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

NL is no limit.

Table C.22. Helping students make connections between their prior knowledge and texts: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Teacher connected big ideas in a text to students' prior knowledge	0	1	0.03	0.46
When students answered questions about the content of a text, the teacher provided specific feedback that helped students arrive at an answer	0	1	0.10	0.43
Teacher connected information about the world to a text the students previously read	0	1	0.05	0.42
Composite item: focusing on the purpose of a text and activating prior knowledge	NL	NL	0.00	0.39
Teacher connected specific details in a text to students' prior knowledge	0	1	0.10	0.36
When students answered questions about the content of a text, the teacher asked students to explain how they figured out their answers	0	1	0.04	0.32
Teacher taught world knowledge related to literary concepts	0	1	0.03	0.31
Teacher connected the general topic of a text to students' prior knowledge	0	1	0.09	0.31

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

NL is no limit.

Table C.23. Teaching students to use other comprehension strategies: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Extent to which teacher provided guidance to students about how to use comprehension strategies on 0-to-4 scale (average across segments)	0	4	0.22	0.91
Specificity of teacher's explanation of how to use comprehension strategies on 0-to-2 scale (average across segments)	0	2	0.09	0.91
Extent to which teacher explained why a comprehension strategy should be used on 0-to-3 scale (average across segments)	0	3	0.11	0.87
Number of comprehension strategies taught on 0-to-2 scale (average across segments)	0	2	0.11	0.86
When students used comprehension strategies, extent of teacher's feedback on 0-to-3 scale (average across segments)	0	3	0.09	0.86
Teacher explained when to use a comprehension strategy	0	1	0.02	0.66

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

Table C.24. Focusing on world knowledge: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Teacher taught world knowledge	0	1	0.47	0.93
Number of minutes (out of 15) in which teacher taught world knowledge (average across segments)	0	12.5	3.59	0.90
Number of pieces of information about the world taught (average across segments)	0	11	2.88	0.90
Students learned world knowledge by reading out loud, discussing questions, writing, drawing, acting, or singing	0	1	0.39	0.87
Composite item: frequency and diversity of world knowledge activities	NL	NL	0.00	0.78
Teacher connected information about the world to something previously learned	0	1	0.18	0.63
Teacher related information about the world to a big idea or theme	0	1	0.14	0.59
Teacher connected information about the world to students' personal experiences	0	1	0.15	0.47
Teacher taught world knowledge about health and science	0	1	0.19	0.45
Teacher taught world knowledge about math	0	1	0.18	0.44
Teacher taught world knowledge about social studies	0	1	0.16	0.42

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

NL is no limit.

Table C.25. Focusing on higher-order thinking: key statistics on the contributing items

Item	Theoretical range of scores		Average score	Factor loading
	Minimum	Maximum		
Number of questions that teacher asked that encouraged students to use higher-order thinking (average across segments)	0	NL	0.77	0.80
Number of minutes (out of 15) in which teacher encouraged students to use higher-order thinking (average across segments)	0	12.5	2.36	0.77
Teacher encouraged higher-order thinking	0	1	0.37	0.76
Number of higher-order questions that asked students to explain their answers or thinking (average across segments)	0	NL	0.35	0.70

Source: Authors' calculations from classroom observation data (N = 4,094 observation sessions).

Note: Observation sessions are the units of analysis. Unless otherwise noted, an item score for an observation session is the fraction of observation segments within that session in which the specified action was observed. Items are ordered from highest to lowest factor loading. Only items with factor loadings of at least 0.30 are listed (even though other items may have loaded onto the factor with lower loadings).

NL is no limit.

Degree of consistency between observers. Generating summary measure scores initially at the level of the observation session provided an additional opportunity to assess interrater reliability—the degree of consistency between observers who observed the same session. As discussed in Appendix B, although each observation session was typically conducted by only one observer, the study team assigned multiple observers to some sessions to check interrater reliability. In Appendix B, we presented the interrater reliability of item scores—specifically, the percentage of item scores (across all observation segments) in which different observers in the same session came to exact agreement.

Here, we present another measure of interrater reliability—the interrater reliability of the summary measure scores, reflecting the degree of consistency between summary measure scores from different observers who observed the same session. Specifically, for the set of observation sessions that had multiple observers, we used the factor analysis model described earlier to generate summary measure scores for each session based on each observer separately. We then calculated an intraclass correlation measure of interrater reliability based on the approach specified by Shrout and Fleiss (1979). Conceptually, intraclass correlations represent the correlation between summary measure scores from different observers in the same session.

Across the 13 summary measures, intraclass correlations ranged from 0.13 to 0.63, for an average of 0.35 (Table C.26). Stated differently, on a typical summary measure, about 35 percent of the variation in the session scores assigned by individual observers represented differences in instructional quality that would be consistently identified by other observers.

Table C.26. Interrater reliability of the session-level summary measures of instructional practices, as measured by intraclass correlations

Instructional practice	Intraclass correlation
Encouraging students' oral language	0.49
Focusing on phonics and grammar during reading	0.54
Engaging students in defining new words during pre-reading	0.52
Engaging students in defining new words during reading	0.16
Engaging students in defining new words during post-reading	0.63
Engaging students in defining new words outside of reading	0.29
Focusing on the meaning of texts during pre-reading	0.13
Focusing on the meaning of texts during reading	0.27
Focusing on the meaning of texts during post-reading	0.17
Helping students make connections between their prior knowledge and texts	0.15
Teaching students to use other comprehension strategies	0.39
Focusing on world knowledge	0.61
Focusing on higher-order thinking	0.19
Average across all practices	**0.35**
Number of observation sessions	**42**
Number of observers	**81**

Source: Authors' calculations from classroom observation data.

Note: For eight sessions in which only a trainer was available to be paired with a regular observer, ratings assigned by both the trainer and regular observer were used.

The interrater reliability of the session-level scores, as reported in Table C.26, did not capture the overall reliability of the final summary measures—the fraction of the variation in these measures that reflected true differences in average instructional quality across classrooms. On the one hand, inconsistencies among observers represented only one of multiple sources of measurement error. As discussed later, even if observers were to agree completely when rating a particular session, the practices observed in that session may not accurately represent what the teacher typically does if his or her practices vary across lessons. This source of measurement error was not reflected in the interrater reliability values of Table C.26. On the other hand, the values in Table C.26 captured the interrater reliability of practice scores from a typical *single* session, but, as discussed in Appendix B, classrooms were actually observed multiple (up to four) times, each time by a different observer. Averaging across multiple observation sessions within a classroom enhanced the reliability of the summary measures; it allowed different observers' biases and errors to offset each other and allowed the study to capture a larger representation of a teacher's practices. For this reason, we averaged across multiple sessions per classroom, as described in the next section. The overall reliability of those classroom-level scores, which we also report in the next section, reflect both the multiple sources of measurement error described above and the reduction of those errors from averaging across sessions and observers.

4. Average summary measures to the classroom level

After obtaining summary measures of instructional practices for each observation session, we averaged the summary measures to the classroom level—specifically, by averaging across the observation sessions for each classroom. (As described in Appendix A, while most classrooms were observed four times, some classrooms were observed in just one, two, or three sessions.) In doing so, we made two adjustments designed to distinguish actual differences in the instructional practices used by teachers from differences due to imperfect measurement of the practices. The

two adjustments, described next, accounted for (1) the time of day in which observation sessions occurred and (2) the limited number of observation sessions per classroom.

Account for time of day. Before averaging the session scores on each summary measure to the classroom level, we adjusted the scores according to whether the session occurred during the morning or afternoon. In general, scores on the summary measures were lower in afternoon sessions. Although most classrooms (68 percent) had equal numbers of morning and afternoon sessions, some (26 percent) had more morning than afternoon sessions, and others (6 percent) had more afternoon than morning sessions. If we had not adjusted the scores, classrooms with more afternoon sessions would tend to have lower average scores on the summary measures. This would cause our estimated relationships between instructional practices and student growth to be too small because the timing of the sessions would affect the instructional practice measures but would not similarly affect student growth (given that students actually experienced both morning and afternoon instruction every day). The adjustment removed the penalty from being observed more frequently in the afternoon.

We carried out the adjustment in two steps. First, to estimate the penalty associated with an afternoon session, we estimated regressions in which the dependent variable consisted of session-level scores on a summary measure and the independent variables consisted of an indicator for whether the session took place in the afternoon and a full set of classroom indicators. The coefficient on the afternoon indicator gave an estimate of the penalty. We performed this step separately for each summary measure and in each grade span. Next, we removed the penalty by subtracting the coefficient from the afternoon session scores.

Account for the limited number of observation sessions per classroom. In each classroom, the instructional practices observed in up to four observation sessions may not have fully represented the teacher's typical practices in that classroom. Because a teacher's practices could vary from one lesson to the next, the limited number of observation sessions could, by chance, have missed the fuller picture of the teacher's instructional practices that would have been revealed with a larger number of observations. Therefore, on each summary measure of an instructional practice, a classroom's average score contained some degree of error as a measure of the true average practice in that classroom.

Failure to account for imperfect measurement of practices would result in smaller relationships (that is, estimated relationships that are biased toward zero) between instructional practices and student growth. This is because, on each summary measure, some of the differences in scores across classrooms were not real differences in teachers' typical practices, but were instead chance differences from having observed some teachers' better-than-usual lessons and other teachers' worse-than-usual lessons. Those chance differences were not associated with student growth. Therefore, adjustments to filter out those chance differences would lead to larger estimates of the relationships between instructional practices and student growth. In fact, the adjustment described below removes the bias in the estimated relationships that would have been caused by measurement error.

We used a statistical approach called empirical Bayes shrinkage (Morris 1983) to account for measurement error in the classroom-level scores on each summary measure. Sullivan (2001) shows how empirical Bayes estimates can be used as independent variables in an ordinary least

squares regression to address measurement error. The approach in Sullivan (2001) is our preferred approach because it provides relationships that account for different amounts of measurement error across classrooms, such as those caused by classrooms with different numbers of observation sessions. The empirical Bayes shrinkage procedure required (1) a *standard error*, an estimate of the amount of measurement error in each classroom-level score and (2) a *prior*, the score that we assumed a classroom would have before seeing the classroom's data, with the assumption being informed by scores from similar classrooms.

The empirical Bayes procedure produced an adjusted estimate of the classroom-level score by combining the actual original score (after accounting for time of day) with the prior in a weighted average. The weight on the original score was larger when it had less measurement error (that is, the standard error was smaller). The weighted average was as follows:

$$(1) \quad P^{EB}_{cgjd} \approx \left(\frac{\hat{\omega}^2_g}{\hat{\omega}^2_g + \hat{\sigma}^2_{cgjd}} \right) P_{cgjd} + \left(1 - \frac{\hat{\omega}^2_g}{\hat{\omega}^2_g + \hat{\sigma}^2_{cgjd}} \right) \theta_{gjd},$$

where P^{EB}_{cgjd} was the empirical Bayes estimate for classroom *c* within grade *g* in school *j* within district *d*, P_{cgjd} was the original score for the same classroom, θ_{gjd} was the prior, $\hat{\sigma}_{cgjd}$ was the standard error of the original classroom-level score, and $\hat{\omega}_g$ was an estimate of the standard deviation of the true practice across classrooms (obtained using the approach described by Morris [1983]), which was constant for all classrooms in a grade span. The term $\hat{\omega}^2_{gjd} / \left(\hat{\omega}^2_{gjd} + \hat{\sigma}^2_{cgjd} \right)$ gives the weight on the original score[12] and must be greater than zero and less than one. Thus, the estimate was always closer to the prior (θ_{gjd}) than the actual score—that is, the estimate "shrank" from the original score. The larger the standard error of the original score—that is, the larger $\hat{\sigma}_{cgjd}$ was—the closer $\hat{\omega}^2_{gjd} / \left(\hat{\omega}^2_{gjd} + \hat{\sigma}^2_{cgjd} \right)$ was to zero and the larger the shrinkage in P_{cgjd}. See Box C.1 for additional discussion of this approach.

[12] In Morris (1983), the empirical Bayes estimate does not exactly equal the precision-weighted average of two values due to a correction for bias in smaller samples. This adjustment decreases the weight on the actual score slightly. For ease of exposition, we have omitted this correction from the description given here.

For each summary measure, we used two steps to estimate the standard error of a classroom-level score. First, we estimated the amount of measurement error in a typical observation session. Conceptually, the more a teacher's practices were inconsistent from one lesson to the next, the more error there was in relying on any one observation session to measure a teacher's average practice. To gauge the degree of inconsistency, we estimated a regression model in which session-level scores were the dependent variable and classroom indicators were the independent variables. The classroom indicators absorbed the variation in scores that was consistent within classrooms, so the variance of the residuals reflected the degree of inconsistency among sessions in the same classroom. We estimated this model separately for each summary measure and for each grade and district combination to obtain distinct measurement error variances. In the second step, we divided the variance estimate by the number of sessions in each classroom. This approach led to lower measurement error variance in classrooms with more sessions, reflecting the additional information about classroom practices provided by each session. The square root of this result was the standard error of the classroom-level score.

Box C.1. Empirical Bayes shrinkage

Empirical Bayes shrinkage reduces the risk that classrooms with small numbers of observation sessions would have high or low scores on the instructional practice measures simply because, by chance, we observed better-than-usual or worse-than-usual lessons. To do so, empirical Bayes shrinkage combines each classroom's original score with a *prior*—an assumption about the classroom's score that can be based on the scores that similar classrooms receive—to produce a final score. A simple prior might be the average of the measure across all classrooms. In this case, empirical Bayes shrinkage would adjust classrooms with fewer sessions more heavily toward the overall average, and the adjustment would be less pronounced for classrooms with more sessions. Essentially, we would rely more heavily on an assumption that a classroom is average if the classroom had fewer sessions.

Thus, empirical Bayes shrinkage assigns each classroom a weighted average of its original score on the practice measure and a prior assumption about the practice in the classroom. The weight on the original score is lower when there is less evidence to support that score, which occurs when the amount of measurement error is higher.

In assigning a prior to each classroom, we followed the approach suggested by Sullivan (2001). Sullivan advocates using a conditional average—an average among classrooms with similar covariate values—as the prior for empirical Bayes shrinkage when including the resulting shrinkage estimates in a subsequent regression analysis with those other covariates. In our subsequent regression analyses to measure the relationships between instructional practices and student growth (described in detail below), the covariates consisted of grade and district indicators to account for the fact that student growth and instructional practices could vary systematically across grades and districts. Therefore, for each summary measure of an instructional practice, a classroom's prior was the average score in the same grade and district. Specifically, we calculated these priors as predictions from a regression of classroom-level scores on grade and district indicators. We estimated these priors and conducted empirical Bayes shrinkage separately for the two grade spans: (1) prekindergarten and kindergarten in the lower grade span, and (2) grades 1 to 3 in the higher grade span. We discuss the rationale for these grade spans later in this appendix.

As discussed earlier, each final classroom-level score was a weighted average of the classroom's original score and the prior. Across classrooms, the average weight on the original

scores measured the *reliability* of the instructional practice measure—the proportion of the variation in the measure that reflected actual differences in teachers' practices (within a grade and district) rather than measurement error. These reliabilities, calculated from the approach specified by Morris (1983), were conceptually analogous to the internal consistency of the classroom-level scores. Reliability was higher to the extent that the session-level scores that were being averaged together showed higher consistency with each other and to the extent that the classroom average was based on more sessions.

Reliability values ranged from 0 to 0.58, with an average value of 0.34, across the 13 summary measures and two grade spans (Table C.27). In other words, about one-third of the variation in practice scores across classrooms represented true differences in the teachers' typical practices rather than instances in which observers just happened to observe atypical instruction. As expected, compared with the reliability values in Table C.27, the weights on specific classrooms' original scores were smaller in classrooms with fewer sessions than average, and larger in classrooms with more sessions than average. We found no meaningful variation in one of the 13 summary measures in the lower grade span, so we excluded this measure—engaging students in defining new words during post-reading—from the analysis in this grade span. This lack of variation may be partially explained by the vocabulary practices having been observed with relatively low frequency (see Tables C.15 through C.18).

Table C.27. Overall reliability of the classroom-level summary measures of instructional practices

Instructional practice	Prekindergarten and kindergarten	Grades 1 to 3	Average across grade spans
Encouraging students' oral language	0.57	0.58	0.58
Focusing on phonics and grammar during reading	0.41	0.28	0.35
Engaging students in defining new words during pre-reading	0.27	0.28	0.28
Engaging students in defining new words during reading	0.44	0.32	0.38
Engaging students in defining new words during post-reading	0.00	0.10	0.05
Engaging students in defining new words outside of reading	0.35	0.37	0.36
Focusing on the meaning of texts during pre-reading	0.26	0.26	0.26
Focusing on the meaning of texts during reading	0.36	0.36	0.36
Focusing on the meaning of texts during post-reading	0.32	0.25	0.29
Helping students make connections between their prior knowledge and texts	0.28	0.25	0.27
Teaching students to use other comprehension strategies	0.37	0.40	0.39
Focusing on world knowledge	0.40	0.42	0.41
Focusing on higher-order thinking	0.39	0.47	0.43
Average across instructional practices	**0.34**	**0.33**	**0.34**

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

As a final step after shrinkage, we scaled the classroom-level measures so that the relationships between instructional practices and student growth could be compared across the

practice measures. To do so, we standardized each of the measures so that the standard deviation of the underlying practice measure—the dispersion in the measure purged of measurement error—was equal to 1. Specifically, we divided each post-shrinkage measure by an estimate of its standard deviation adjusted for measurement error. We estimated the adjusted standard deviation $\hat{\omega}_g$ separately by grade span using the approach described by Morris (1983).

Interpretation of the reliability of the summary measures. When measuring any behavior, the reliability of the measure critically shapes a study's ability to detect true relationships between the behavior and the outcomes of interest. The reliability values shown in Table C.27 indicate that the classroom-level scores on the summary measures of instructional practices had limited reliability. However, in what follows, we show that these reliability values were similar to those obtained in prior research based on a different and widely used observation tool. In fact, when designing the study, we drew upon this research to anticipate the reliability values shown in Table C.27 and structured key aspects of the sample and analysis so that the limited reliability would not undermine either the validity or precision of the study's estimates.

Prior research has shown that observational measures of instructional quality are vulnerable to many sources of measurement error (Raudenbush et al. 2011). When an observer assigns a score on instructional quality after observing a short segment of instruction, the score may deviate from the teacher's average instructional quality for various reasons. For example, the observer's conclusion might be atypical of what other observers would have concluded by observing the same classroom (observer and observer-by-classroom effects); the teacher's instructional practices on the day of the observation might be atypical of those that the teacher normally does (day and day-by-classroom effects); and the teacher's practices during a short observed segment may not even be typical of those that she did throughout the same day (segment effects).

Based on data from a widely used observation instrument, Raudenbush et al. (2011) showed that these sources of measurement error are collectively large in magnitude. Their data came from a large-scale study (Pianta et al. 2005) in which observers rated 240 preschool classrooms across six states using a well-known instrument, the Classroom Assessment Scoring System (CLASS; Pianta et al. 2006). For the component of the CLASS measuring instructional climate, Raudenbush and colleagues found that only 10 percent of the total variation in segment scores captured differences in average instructional quality across classrooms. The remaining variation reflected observer and observer-by-classroom effects (37 percent), day and day-by-classroom effects (19 percent), and segment effects (34 percent). Rating each classroom based on scores averaged across multiple segments per day, multiple days, and multiple observers would reduce the magnitude of these errors.

At the design phase of this study, we used evidence from Raudenbush et al. (2011) to project the reliability of our instructional practice measures based on the study's actual design for the numbers of segments, observation sessions (analogous to "days"), and observers per classroom. In other words, we assumed that segment scores from the OLLI would have exactly the same sources and magnitude of measurement error as the CLASS, and we projected what the magnitude of these errors would be after averaging across six segments per session and four sessions per classroom, with each session observed by a different observer. We projected the

reliability of this study's summary measures of instructional practices to be 0.39, which was very similar to the average reliability that we actually achieved (0.34), as reported in Table C.27.

Imperfect reliability of instructional practice measures would typically lead to either bias or imprecision in estimating relationships with student growth, but neither problem occurred in this study because we took the reliability projections into account when designing the sample and analysis. As discussed above, the potential concern with bias is that because instructional practice measures ultimately serve as independent variables in regressions for predicting student growth, unreliability in those variables can lead to downward bias in the estimated regression coefficients. However, the use of empirical Bayes shrinkage to adjust independent variables for their degree of unreliability removes this downward bias (Sullivan 2001). Even though this procedure removes bias, it effectively reduces the variation in the instructional practice measures and, consequently, diminishes the precision with which we can estimate their relationships with student growth. However, because this study included a large number of classrooms, the estimated relationships still demonstrated adequate levels of precision. In Section C of this appendix, we show that this study had enough precision to detect the full range of relationships that might be of interest for informing future impact evaluations.

B. Measuring teachers' contributions to student growth

After creating summary measures of instructional practices, the next step was to measure teachers' contributions to student growth. (Because each teacher taught only one classroom in the study, teachers' contributions are equivalent to classrooms' contributions, and we refer to teachers and classrooms interchangeably.) To measure the contribution of each classroom teacher to student growth, we estimated the following regression model, separately for each grade span and student outcome:

$$(2)\ Y^{s}_{icgjd} = \lambda Y^{f}_{i} + \delta X_{i} + \mu_{cgjd} + \varepsilon_{icgjd},$$

where Y^{s}_{icgjd} was the spring score on the outcome for student *i* in classroom *c* within grade *g* in school *j* within district *d*; Y^{f}_{i} was the student's fall score on the same outcome; X_{i} was a set of other covariates, including scores on other fall assessments and demographic characteristics; μ_{cgjd} was a classroom fixed effect (that is, a set of coefficients on binary indicators for each classroom) that we estimated; ε_{icgjd} was a random error term; and λ and δ were parameters that we estimated.

The key estimates from equation (2) were the estimated classroom fixed effects, $\hat{\mu}_{cgjd}$, which measured the contribution of each classroom's teacher to student growth. In essence, equation (2) predicted each student's spring score based on all available fall test scores and background characteristics, and each teacher's contribution, $\hat{\mu}_{cgjd}$, was the average difference between the actual and predicted scores of his or her students.

Including the binary indicators for each classroom when predicting students' spring scores in the regression provides better predictions of students' spring scores and, consequently, better measures of students' growth. By including these classroom indicators in equation (2), we compared the characteristics of students within the same classrooms when estimating the relationships between background characteristics—including fall scores—and spring scores. Otherwise, we would have risked confounding the relationships between characteristics and spring scores with how teachers were sorted to classrooms and schools (Guarino et al. 2014). For example, students with higher fall scores might be taught by more effective teachers. If so, their predicted spring scores would be too high, reflecting not only their higher fall scores, but also that they had more effective teachers. In this example, the measured growth of students in these classrooms would be too low, leading to potential bias in the estimated relationships between practices and student growth described in Section C of this appendix. Thus, our approach separates the influence of student background characteristics (over which teachers have no control) from the influence of classroom-level factors such as instructional quality (over which teachers have control).

The demographic variables were indicators for females; whether a student's home language was English, Spanish, or another language (the excluded category); English as a second language status; special education status; and whether the student was old for his or her grade. All test score and demographic variables in the regression were centered at the mean of the estimation sample.

For each outcome, we estimated equation (2) using all students who had both fall and spring scores on that outcome. For all other covariates, we set missing values to zero and included indicator variables in the regression for whether a student was missing the original value of the covariate. We excluded a constant term from the regression so that indicators for each classroom could be included. The regression was weighted using analysis weights that accounted for the study's sampling design and nonresponse (see Appendix A). For this step, we used Huber-White robust standard errors that accounted for arbitrary differences in the magnitude of regression errors across classrooms (Huber 1967; White 1980).

We calculated teachers' contributions to student growth on three outcomes in the lower grade span and four in the upper grade span. Teachers' contributions may be similar across some or all of these outcomes. If so, the relationships between instructional practices and student growth may also be similar. To assess this possibility, we estimated correlations between teachers' contributions to student growth on different outcomes. These correlations were not high, ranging from 0.13 to 0.38 (Table C.28). In other words, teachers who made the strongest contributions to one outcome were not always the same as those making strong contributions to other outcomes. Therefore, instructional practices that are associated with teachers' contributions to a particular outcome may not necessarily be associated with their contributions to other outcomes.

Table C.28. Correlations between teachers' contributions to student growth across outcomes

Student outcome	Basic language skills	Background knowledge	Listening comprehension
Prekindergarten and kindergarten			
Background knowledge	0.38		
Listening comprehension	0.35	0.37	
Grades 1 to 3			
Background knowledge	0.24		
Listening comprehension	0.26	0.36	
Reading comprehension	0.25	n.a.[a]	0.13

Source: Authors' calculations using data from the fall and spring tests administered by the study team.

Note: In the lower grades, outcomes were measured in 378 classrooms. In the upper grades, background knowledge was measured in 220 classrooms in grade 1, reading comprehension was measured in 435 classrooms in grades 2 and 3, and the remaining two outcomes were measured in 657 classrooms in grades 1 to 3.

[a] No teachers had growth estimates in both reading comprehension and background knowledge because the outcomes were measured in different grades.

C. Assessing the relationships between instructional practices and teachers' contributions to student growth

1. Main analyses

To examine the relationship between instructional practices and student growth, we estimated the following regression model:

$$(3)\ \hat{\mu}_{cgjd} = \alpha + \beta P^{EB}_{cgjd} + \upsilon_g + \pi_d + \varphi_{cgjd},$$

where $\hat{\mu}_{cgjd}$ was the estimate of a classroom teacher's contribution to student growth (obtained from equation [2], discussed earlier); P^{EB}_{cgjd} was the classroom's score on the summary measure of a particular instructional practice (obtained from equation [1], discussed earlier); υ_g was a grade fixed effect; π_d was a district fixed effect; φ_{cgjd} was a random error term; and α and β were parameters that we estimated. In our main analyses, we estimated equation (3) separately for each student outcome, instructional practice, and grade span. Because schools were the largest units that we randomly sampled for the study, we used cluster-robust standard errors that accounted for arbitrary correlation of the regression errors within schools (Liang and Zeger 1986). Each classroom was weighted by the sum of students' analysis weights in that classroom.

The key coefficient in equation (3), β, represented the relationship between the instructional practice and student growth. Specifically, it represented the change in student growth on the outcome, measured in student-level standard deviations of fall scores, that was associated with a one standard deviation increase in a classroom's score on the instructional practice. For example, a coefficient of 0.1 meant that when comparing classrooms at the 50th and 84th percentiles of an instructional practice measure (a difference of one standard deviation),

growth of the average student in the latter classroom would be predicted to be higher by 4 percentile points (a difference of 0.1 standard deviations).

To provide additional context for understanding the magnitude of our results, we also presented the results in terms of the proportion of variation in student growth across classrooms that was explained by the instructional practice measure (as opposed to other factors, such as other unmeasured practices). Calculating this proportion required several steps, which we performed for each practice, outcome, and grade span. First, we removed the variation in the classroom growth estimates due to grade and district by calculating the residuals from a version of equation (3) without the practice measure. Next, we measured the percentage of the residual variance explained by the practice measure. This value is the R-squared from a version of equation (3) that used the residuals in place of the classroom growth estimates as the dependent variable. However, this percentage was initially too small because the residuals included sampling error due to small numbers of students per classroom. To address this concern, we made a final adjustment. We estimated the reliability of the classroom growth estimates using the same approach described by Morris (1983) that we used earlier to estimate the reliability of the instructional practice measures. We then divided the R-squared by the estimated reliability of the classroom growth estimates.

We estimated separate models for each grade span to account for the possibility that instructional practices could have different relationships with language development for younger and older children. We grouped pre-kindergarten and kindergarten separately from grades 1 to 3 because an initial descriptive analysis of the items that underlie the instructional practice measures indicated that the practices used in grade 1 were more similar to the practices used in grades 2 and 3 than to those in the lower grades.

We also considered, but rejected, grouping kindergarten with grade 1. U.S. kindergartens have been undergoing a major transition over the past decade (Bassok et al. 2016). Nevertheless, although many instructional practices in kindergartens are becoming more like those in first-grade classrooms, there are still many factors that distinguish kindergarten from grade 1, including the incidence of half-day versus full-day enrollment (Child Trends Databank 2015), the relative focus on emergent literacy skills versus conventional literacy (National Early Literacy Panel 2008), and the proportion of play to academic work (Bassock et al. 2016). Moreover, grouping kindergarten with grade 1 would have left only a single grade (prekindergarten) in the lower grade span, for which we would not have adequate precision to detect relationships between practices and growth.

The estimated relationships between instructional practices and student growth were more precise when based on more classrooms, and when there were more measurable differences across classrooms in instructional practices. The number of classrooms varied across outcomes from 220 to 657. As noted above, the reliability of the instructional practice scores—which indicated the extent to which variation in these measures reflected true differences across classrooms—varied from 0 to 0.58 (Table C.27). To gauge how these factors could affect precision, we calculated the minimum detectable relationships (MDR) between practices and student growth—that is, the smallest true relationships for which our analysis would find a statistically significant result with 80 percent probability.

The study had considerable precision for estimating relationships between practices and growth. The MDR values—expressed as the increase in student scores, in standard deviation units, associated with a one standard deviation increase in an instructional practice across classrooms—ranged from 0.06 to 0.12, depending on the instructional practice. Smaller MDR values, which indicated better precision, tended to be attained for more reliable instructional practices. For example, the most reliably measured instructional practice in the upper grade span, encouraging students' oral language, had an MDR of 0.06, whereas engaging students in defining new words during post-reading, the least reliably measured practice, had an MDR of 0.12. Nevertheless, all of these MDR values were smaller than the effect sizes that previous rigorous evaluations of early literacy programs were designed to detect. For example, the Early Reading First study was designed to detect an effect size no smaller than 0.30 (Jackson et al. 2007). Therefore, this study could detect the full range of relationships that might be of interest for informing future impact evaluations.

In summary, our analysis of relationships between instructional practices and student growth took a two-step approach in which we first estimated teachers' contributions to student growth (equation [2]) and then assessed how those contributions related to instructional practices (equation [3]).

Before conducting the analyses, we considered, but did not adopt, an alternative one-step approach that would entail estimating a single regression model for the relationship between each instructional practice (the independent variable of interest) and students' spring scores (the dependent variable) while controlling for students' fall scores and other characteristics. Such a model could be estimated with methods that account for the clustering of students within classrooms—for instance, ordinary least squares with cluster-robust standard errors or multilevel mixed-effects models. This alternative approach would have been similar to our two-step approach in a number of ways. Both approaches make use of all available student-level data (on fall and spring test scores) and classroom-level data (on instructional practices). The two-step approach uses student-level data in the initial step and classroom-level data in the subsequent step, whereas the one-step approach uses all of this data at once. Also, both approaches hold constant individual students' fall scores when comparing outcomes across classrooms with different instructional practices. Fall scores are a covariate in the initial step of the two-step approach and in the single regression of the one-step approach.

We chose the two-step approach because it was better able than the one-step approach at isolating *teachers' contributions* to student growth—rather than student growth more generally—as the outcome to be examined. As discussed in Chapter II, we focused on identifying differences across classrooms in teachers' contributions to student growth because those were the differences that could result from teachers' instructional practices. Earlier, in Section B of this appendix, we showed that the key step in identifying teachers' contributions to student growth was to compare students' spring outcomes with the outcomes they would be predicted to have based on their fall scores (and other characteristics), had they been taught by the average teacher. To obtain this prediction, we must identify the correct relationship between fall scores and spring scores that would occur were students to be taught by the average teacher. As discussed in Section B, the advantage of the two-step approach is that it allowed us to control for classroom indicators when estimating the relationship between fall and spring scores (in the first step, equation [2]). Essentially, including the classroom indicators held teacher effectiveness

constant when estimating the relationship between fall and spring scores. In contrast, classroom indicators could not be included in the one-step approach because they would absorb all of the variation in the instructional practice measures. Therefore, the one-step approach cannot hold teacher effectiveness constant when estimating the fall-to-spring test score relationship. If we could not hold teacher effectiveness constant, and if more effective (or less effective) teachers tended to be assigned to classrooms with higher baseline achievement, then we would overestimate (or underestimate) the relationship between fall and spring scores. Incorrect estimates of this relationship would, in turn, lead to erroneous estimates of teachers' contributions to student growth as the outcome being examined.

2. Subgroup analyses

This study also sought to identify instructional practices that were potentially promising for particular types of students. To do so, we examined relationships between instructional practices and student growth separately within each of six subgroups: (1) students whose home language was English, (2) students whose home language was not English, (3) students with high baseline test scores ("high achievers"), (4) students with low baseline test scores ("low achievers"), (5) boys, and (6) girls. For each subgroup, we first estimated equation (2) using only the students in that subgroup, producing estimates of individual teachers' contributions to the growth of students in the specified subgroup. We then used those estimates of teachers' contributions as the outcome in equation (3). The resulting estimates of equation (3) captured the relationships between instructional practices and teachers' contributions to the growth of students in the specified subgroup.

Because home language and gender are categorical characteristics, we formed subgroups along those dimensions by classifying students directly into the home language and gender categories. In contrast, students' baseline achievement varied along a continuum. Therefore, we needed to carve out categories from this continuous characteristic so that we would have well-defined subsamples of students on which we would estimate relationships between practices and growth. The remainder of this section explains how we defined the categories of high achievers and low achievers.

The definitions of high and low achievers were specific to the outcome being examined. Separately for each of the four outcomes—basic language skills, background knowledge, listening comprehension, and reading comprehension—we defined low and high achievers to be students whose fall score on the same assessment as the outcome was in the bottom 40 percent and top 40 percent, respectively, of the analysis sample for that outcome. When defining these subgroups, we excluded the middle 20 percent of students so that the resulting subgroups would be somewhat more extreme in their baseline achievement than they would have been had the middle achievers been included. If certain practices were potentially promising for very high or very low achievers, creating somewhat more extreme subgroups would increase the likelihood of identifying such practices. Given the study's exploratory aim, this approach was consistent with our overall analytic strategy, discussed in Chapter II, of trying to find as many potentially promising practices as warranted by the evidence.

Although creating subgroups of high and low achievers entailed carving categories out of a continuous characteristic, this analysis was not vulnerable to the oft-cited shortcomings of converting continuous variables into categorical variables. As discussed in many prior studies,

when analyzing the relationships between two continuous variables, converting one or both of those variables into categorical measures artificially decreases the size of the estimated correlations and lowers statistical power for detecting a true relationship (see, for example, MacCallum et al. [2002] and Royston et al. [2006]). However, in our analysis, we were not examining the direct relationship between baseline achievement and outcomes, nor were we testing whether the relationship between practices and outcomes differed by students' baseline achievement. Instead, we were defining categories of baseline achievement to create subgroups within which we *separately* examined the relationships between practices and outcomes. None of the methodological objections to creating categories out of continuous variables applied to this scenario.

This page has been left blank for double-sided copying.

APPENDIX D: ADDITIONAL RESULTS

In this appendix, we summarize the results of additional analyses that relate to the study's goal of identifying relationships between instructional practices and students' growth in language and comprehension. These analyses include the following:

- Assessing the magnitudes of the relationships between instructional practices and student growth and determining whether these relationships are statistically significant according to different significance criteria
- Exploring relationships that include all practices simultaneously to account for how teachers might use multiple practices
- Exploring relationships that account for the prerequisite actions that must occur before some types of practices are implemented
- Analyzing relationships within student subgroups

A. Magnitudes and statistical significance of the relationships between instructional practices and student growth in language and comprehension

In Chapter III of the main report, we summarized the direction and statistical significance of the relationships between instructional practices and student growth in language and comprehension. Here, we provide detailed results on the magnitudes of those relationships. We assessed the magnitude of the relationship between each instructional practice and student growth in two ways (see Appendix C for technical details). First, we measured the relationship size—the change in student test scores, measured in student-level standard deviations, that was associated with a one standard deviation increase in the instructional practice across classrooms. Second, we measured the percentage of variation in growth across classrooms that was explained by each instructional practice. This second approach was designed to provide additional context for the magnitude of the relationships. A small relationship measured in student-level standard deviations could represent a large percentage of the variation in growth if there is limited variation in growth in the language or comprehension outcome across classrooms. In both approaches, the magnitudes of the relationships between practices and growth were expressed on a common scale across practices, outcomes, and grade spans.

In this appendix, we also provide additional information about the statistical significance of the relationships. In the main report (Tables III.1 and III.2), we identified relationships as statistically significant when the *p*-value was less than 0.10. Although significance at the 0.05 level is a more typical threshold for statistical significance, we chose a more generous threshold to reduce the risk of failing to identify a practice that deserved further study. Here, we indicate whether a relationship was significant only at the 0.10 level, at the 0.05 level, or at the 0.05 level after a multiple comparisons adjustment. We applied the Benjamini-Hochberg adjustment (Benjamini and Hochberg 1995) for multiple comparisons within each instructional practice, across outcomes and grade levels.

Individual practices explained between 0 and 14 percent of the variation in growth. Focusing on higher-order thinking explained 14 percent of the variation in background knowledge growth

in the upper grades, the largest proportion of any practice and student growth outcome. The remaining variation might be explained by other instructional practices, unmeasured aspects of classrooms and instruction (such as access to resources), or unmeasured differences in student characteristics across classrooms. Tables D.1 to D.5 report these magnitudes and the *p*-values of the relationship sizes.

Using the results in Tables D.1 to D.5, we also explored which practices that were identified as potentially promising continued to be identified as such under more stringent requirements for statistical significance. As explained in Chapter II, this study regards a practice to be potentially promising within a grade span if it had a positive, statistically significant relationship with at least one outcome and no negative, statistically significant relationships with any other outcome in that grade span. When identifying potentially promising practices, we always considered a negative finding to be statistically significant using the 0.10 level, even when using the more stringent requirements for positive findings. Otherwise, it would be possible for a practice that was not identified as potentially promising using the main approach to be identified as potentially promising using the more stringent approach.

As expected, we identified fewer practices as potentially promising when using more stringent requirements for statistical significance compared to the results from the main approach. In the lower grades, four of the five practices identified as potentially promising using the main approach remained so when requiring significance at the 0.05 level instead of the 0.10 level. Focusing on world knowledge is no longer a potentially promising practice when using the more stringent threshold. Requiring significance at the 0.05 level did not change the five potentially promising practices in the upper grades. When also adjusting for multiple comparisons, we identified only one practice as potentially promising in the lower grades: focusing on the meaning of texts during pre-reading. In the upper grades, engaging students in defining new words during post-reading and focusing on higher-order thinking are potentially promising when adjusting for multiple comparisons. These results are summarized in the last two columns of Tables D.15 and D.16.

Table D.1. Relationships between instructional practices that encourage students' oral language or focus on phonics and grammar during reading and student growth in language and comprehension

	Prekindergarten and kindergarten			Grades 1 to 3		
Student outcome	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)
Association between encouraging students' oral language and						
Basic language skills	0.00	0.80	<0.1	0.02	0.24	1.3
Background knowledge	0.00	0.95	<0.1	0.08**	0.02	8.9
Listening comprehension	-0.02	0.55	0.6	0.01	0.70	n.r.[a]
Reading comprehension				0.00	0.99	<0.1
Association between focusing on phonics and grammar during reading and						
Basic language skills	0.00	0.88	0.1	0.01	0.53	0.1
Background knowledge	-0.04	0.33	0.8	0.07	0.23	4.7
Listening comprehension	-0.09**	0.02	8.3	0.03	0.23	n.r.[a]
Reading comprehension				0.03	0.47	0.9
Number of classrooms	**378**			**220–657**		

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Variation in growth is the variance of student growth across classrooms, excluding variation due to measurement error. Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

++Significantly different from zero at the .05 level after an adjustment for multiple comparisons, two-tailed test.

n.r. = not reported

[a] The percentage of the variation in growth explained by the practice is suppressed for listening comprehension in grades 1 to 3 because the variance in student growth across classrooms was too small after accounting for measurement error. Using this small number in the denominator would produce unstable measures of the percentage of variation explained.

Table D.2. Relationships between instructional practices that engage students in defining new words and student growth in language and comprehension

	Prekindergarten and kindergarten			Grades 1 to 3		
Student outcome	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)
Association between engaging students in defining new words during pre-reading and						
Basic language skills	-0.04	0.17	1.7	-0.05	0.17	2.4
Background knowledge	-0.12++	<0.01	7.2	-0.04	0.33	1.2
Listening comprehension	-0.06**	0.02	4.1	0.03	0.20	n.r.[a]
Reading comprehension				0.02	0.54	<0.1
Association between engaging students in defining new words during reading and						
Basic language skills	0.06**	0.04	6.1	-0.03	0.12	1.9
Background knowledge	-0.05	0.24	1.5	0.09**	0.01	7.0
Listening comprehension	0.01	0.81	<0.1	0.02	0.58	n.r.[a]
Reading comprehension				-0.03*	0.07	1.0
Association between engaging students in defining new words during post-reading and[b]						
Basic language skills				0.03	0.56	0.9
Background knowledge				0.04	0.57	0.7
Listening comprehension				0.06	0.15	n.r.[a]
Reading comprehension				0.15++	<0.01	6.1
Association between engaging students in defining new words outside of reading and						
Basic language skills	-0.05	0.17	3.7	0.03	0.10	3.6
Background knowledge	0.00	0.98	<0.1	0.05	0.21	1.7
Listening comprehension	0.02	0.62	0.3	-0.01	0.67	n.r.[a]
Reading comprehension				-0.01	0.59	0.4
Number of classrooms	**378**			**220–657**		

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Variation in growth is the variance of student growth across classrooms, excluding variation due to measurement error. Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

[a] The percentage of the variation in growth explained by the practice is suppressed for listening comprehension in grades 1 to 3 because the variance in student growth across classrooms was too small after accounting for measurement error. Using this small number in the denominator would produce unstable measures of the percentage of variation explained.

[b]The summary measure of engaging students in defining new words during post-reading did not vary across prekindergarten and kindergarten classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

++Significantly different from zero at the .05 level after an adjustment for multiple comparisons, two-tailed test.

n.r. = not reported

Table D.3. Relationships between instructional practices that focus on the meaning of texts and student growth in language and comprehension

	Prekindergarten and kindergarten			Grades 1 to 3		
Student outcome	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)
Association between focusing on the meaning of texts during pre-reading and						
Basic language skills	0.09++	<0.01	9.3	-0.02	0.39	0.7
Background knowledge	0.00	0.91	<0.1	-0.06	0.27	2.7
Listening comprehension	0.03	0.34	0.6	0.01	0.59	n.r.[a]
Reading comprehension				-0.03	0.44	0.4
Association between focusing on the meaning of texts during reading and						
Basic language skills	0.03	0.41	0.6	-0.04**	0.04	4.5
Background knowledge	-0.10**	0.01	3.3	0.08**	0.04	5.5
Listening comprehension	-0.03	0.41	0.6	0.00	0.87	n.r.[a]
Reading comprehension				-0.01	0.76	<0.1
Association between focusing on the meaning of texts during post-reading and						
Basic language skills	0.02	0.42	1.2	0.00	0.92	0.2
Background knowledge	-0.02	0.64	<0.1	0.05	0.23	1.5
Listening comprehension	0.01	0.83	<0.1	0.02	0.54	n.r.[a]
Reading comprehension				-0.01	0.89	<0.1
Number of classrooms	**378**			**220–657**		

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Variation in growth is the variance of student growth across classrooms, excluding variation due to measurement error. Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

++Significantly different from zero at the .05 level after an adjustment for multiple comparisons, two-tailed test.

n.r. = not reported

[a] The percentage of the variation in growth explained by the practice is suppressed for listening comprehension in grades 1 to 3 because the variance in student growth across classrooms was too small after accounting for measurement error. Using this small number in the denominator would produce unstable measures of the percentage of variation explained.

Table D.4. Relationships between instructional practices that help students use comprehension strategies and student growth in language and comprehension

	Prekindergarten and kindergarten			Grades 1 to 3		
Student outcome	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)
Association between helping students make connections between their prior knowledge and texts and						
Basic language skills	0.04	0.29	1.3	0.03	0.18	1.1
Background knowledge	0.02	0.50	0.2	0.07*	0.10	2.4
Listening comprehension	0.08**	0.01	4.4	0.03	0.36	n.r.[a]
Reading comprehension				0.07**	0.05	2.5
Association between teaching students to use other comprehension strategies and						
Basic language skills	0.00	0.91	<0.1	0.01	0.59	0.7
Background knowledge	-0.02	0.70	0.1	-0.01	0.81	<0.1
Listening comprehension	-0.02	0.63	0.2	0.05**	0.03	n.r.[a]
Reading comprehension				0.03	0.41	0.7
Number of classrooms	**378**			**220–657**		

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Variation in growth is the variance of student growth across classrooms, excluding variation due to measurement error. Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

++Significantly different from zero at the .05 level after an adjustment for multiple comparisons, two-tailed test.

n.r. = not reported

[a] The percentage of the variation in growth explained by the practice is suppressed for listening comprehension in grades 1 to 3 because the variance in student growth across classrooms was too small after accounting for measurement error. Using this small number in the denominator would produce unstable measures of the percentage of variation explained.

Table D.5. Relationships between instructional practices that focus on world knowledge or higher-order thinking and student growth in language and comprehension

	Prekindergarten and kindergarten			Grades 1 to 3		
Student outcome	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)	Relationship size	*p*-value	Variation in growth explained by the practice (percentage)
Association between focusing on world knowledge and						
Basic language skills	0.05*	0.08	2.7	-0.05**	<0.01	4.4
Background knowledge	0.00	0.97	<0.1	0.06	0.19	3.1
Listening comprehension	0.01	0.49	0.2	-0.02	0.30	n.r.[a]
Reading comprehension				-0.01	0.61	0.3
Association between focusing on higher-order thinking and						
Basic language skills	0.04**	0.05	2.7	0.01	0.80	0.6
Background knowledge	0.03	0.32	0.5	0.10++	<0.01	14.4
Listening comprehension	0.04	0.27	2.1	-0.02	0.43	n.r.[a]
Reading comprehension				0.00	0.92	<0.1
Number of classrooms	**378**			**220–657**		

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Variation in growth is the variance of student growth across classrooms, excluding variation due to measurement error. Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

++Significantly different from zero at the .05 level after an adjustment for multiple comparisons, two-tailed test.

n.r. = not reported

[a] The percentage of the variation in growth explained by the practice is suppressed for listening comprehension in grades 1 to 3 because the variance in student growth across classrooms was too small after accounting for measurement error. Using this small number in the denominator would produce unstable measures of the percentage of variation explained.

B. Relationships that account for how teachers might use multiple practices

Teachers who differ in how frequently they use one practice may also differ in their frequency of using other practices. The main results in Chapter III of the report documented relationships between each practice and student growth without accounting for the possibility that teachers use multiple practices. Therefore, the association between a practice and student growth may have reflected the influence of other practices.

To address this concern, we also conducted analyses that included all 13 practices simultaneously—that is, we measured relationships for each instructional practice that accounted for all other practices. The results are reported in Tables D.6 to D.9, with one table for each student growth outcome. These analyses tended to have limited statistical precision. For example, when two practices tended to occur together frequently, isolating the relationship between each practice and student growth needed to rely on the rare cases in which some teachers frequently used one practice but not the other.

In general, the results were similar to the main results from Chapter III that did not account for other practices, but we also found some differences. Seven of the 13 positive and significant relationships from our main analysis were not significant when accounting for all other practices, even though in many cases the relationship sizes did not change substantially. Also, one relationship that was positive but insignificant in our main analysis became significant. Based on these results that account for other practices, three practices in the lower grades and one in the upper grades are not identified as potentially promising in this analysis, but were in our main analysis. Nevertheless, we caution that this approach, due its limited precision, was more likely than the main approach to overlook promising practices. The practices identified as potentially promising when accounting for other practices are listed in the second column in each of Tables D.15 and D.16.

When accounting for all practices simultaneously, the practices explained a larger percentage of the variation in growth across classrooms. Although larger, the percentage of variation explained by the practices when all 13 were included simultaneously was approximately the same or less than the sum of the percentages when including each of the 13 practices individually. This finding indicates that some teachers did use multiple practices. Whereas individual practices explained between 0 and 14 percent of the variation in growth, all 13 practices combined explained between 12 and 32 percent, depending on the outcome and grade span.

Table D.6. Relationships between instructional practices and student growth in basic language skills when accounting for other practices

	Prekindergarten and kindergarten		Grades 1 to 3	
Instructional practice	Relationship size	*p*-value	Relationship size	*p*-value
Encouraging students' oral language	-0.03	0.22	0.03	0.16
Focusing on phonics and grammar during reading	-0.01	0.64	0.01	0.66
Engaging students in defining new words during pre-reading	-0.05*	0.07	-0.07**	0.04
Engaging students in defining new words during reading	0.06	0.12	-0.03	0.23
Engaging students in defining new words during post-reading[a]			0.04	0.38
Engaging students in defining new words outside of reading	-0.05	0.21	0.03	0.13
Focusing on the meaning of texts during pre-reading	0.09**	0.03	0.01	0.69
Focusing on the meaning of texts during reading	-0.03	0.42	-0.04	0.14
Focusing on the meaning of texts during post-reading	0.02	0.52	-0.02	0.42
Helping students make connections between their prior knowledge and texts	0.01	0.72	0.06**	0.03
Teaching students to use other comprehension strategies	-0.02	0.62	0.03	0.19
Focusing on world knowledge	0.03	0.40	-0.06***	<0.01
Focusing on higher-order thinking	0.03	0.15	0.01	0.50
Variation in growth explained by all instructional practices (percentage)	24.8		27.4	
Number of classrooms	**378**		**657**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

[a]The summary measure of engaging students in defining new words during post-reading did not vary across prekindergarten and kindergarten classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.7. Relationships between instructional practices and student growth in background knowledge when accounting for other practices

	Prekindergarten and kindergarten		Grade 1	
Instructional practice	Relationship size	*p*-value	Relationship size	*p*-value
Encouraging students' oral language	0.03	0.45	0.06	0.11
Focusing on phonics and grammar during reading	-0.02	0.64	0.01	0.80
Engaging students in defining new words during pre-reading	-0.12***	<0.01	-0.08*	0.08
Engaging students in defining new words during reading	-0.03	0.49	0.05	0.30
Engaging students in defining new words during post-reading			0.02	0.75
Engaging students in defining new words outside of reading	0.00	1.00	0.03	0.50
Focusing on the meaning of texts during pre-reading	0.06	0.36	-0.11	0.10
Focusing on the meaning of texts during reading	-0.11**	0.04	0.02	0.59
Focusing on the meaning of texts during post-reading	0.02	0.68	-0.01	0.84
Helping students make connections between their prior knowledge and texts	0.04	0.25	0.04	0.36
Teaching students to use other comprehension strategies	0.03	0.58	-0.02	0.71
Focusing on world knowledge	0.02	0.66	-0.01	0.78
Focusing on higher-order thinking	0.02	0.52	0.07*	0.10
Variation in growth explained by all instructional practices (percentage)	12.5		32.8	
Number of classrooms	**378**		**220**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Variation in growth is the variance of student growth across classrooms, excluding variation due to measurement error.

[a]The summary measure of engaging students in defining new words during post-reading did not vary across prekindergarten and kindergarten classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.8. Relationships between instructional practices and student growth in listening comprehension when accounting for other practices

Instructional practice	Prekindergarten and kindergarten		Grades 1 to 3	
	Relationship size	*p*-value	Relationship size	*p*-value
Encouraging students' oral language	-0.02	0.70	0.01	0.72
Focusing on phonics and grammar during reading	-0.11**	0.02	0.03	0.38
Engaging students in defining new words during pre-reading	-0.08***	<0.01	0.02	0.44
Engaging students in defining new words during reading	0.03	0.58	0.01	0.77
Engaging students in defining new words during post-reading[a]			0.04	0.42
Engaging students in defining new words outside of reading	0.04	0.35	-0.01	0.67
Focusing on the meaning of texts during pre-reading	0.06	0.25	-0.01	0.65
Focusing on the meaning of texts during reading	-0.01	0.74	-0.03	0.44
Focusing on the meaning of texts during post-reading	0.03	0.35	0.00	0.97
Helping students make connections between their prior knowledge and texts	0.08**	0.02	0.02	0.57
Teaching students to use other comprehension strategies	0.01	0.85	0.05**	0.02
Focusing on world knowledge	-0.02	0.56	-0.02	0.34
Focusing on higher-order thinking	0.04	0.39	-0.03	0.31
Variation in growth explained by all instructional practices (percentage)	24.3		n.r.[b]	
Number of classrooms	**378**		**657**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Variation in growth is the variance of student growth across classrooms, excluding variation due to measurement error.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

n.r. = not reported

[a]The summary measure of engaging students in defining new words during post-reading did not vary across prekindergarten and kindergarten classrooms.

[b] The percentage of the variation in growth explained by the practice is suppressed for listening comprehension in grades 1 to 3 because the variance in student growth across classrooms was too small after accounting for measurement error. Using this small number in the denominator would produce unstable measures of the percentage of variation explained.

Table D.9. Relationships between instructional practices and student growth in reading comprehension when accounting for other practices

	Grades 2 to 3	
Instructional practice	Relationship size	*p*-value
Encouraging students' oral language	-0.01	0.89
Focusing on phonics and grammar during reading	0.04	0.37
Engaging students in defining new words during pre-reading	0.01	0.70
Engaging students in defining new words during reading	-0.05*	0.07
Engaging students in defining new words during post-reading	0.16***	<0.01
Engaging students in defining new words outside of reading	-0.01	0.84
Focusing on the meaning of texts during pre-reading	-0.05	0.16
Focusing on the meaning of texts during reading	0.00	0.95
Focusing on the meaning of texts during post-reading	-0.05	0.20
Helping students make connections between their prior knowledge and texts	0.08**	0.05
Teaching students to use other comprehension strategies	0.03	0.37
Focusing on world knowledge	0.00	0.96
Focusing on higher-order thinking	0.01	0.82
Variation in growth explained by all instructional practices (percentage)	14.1	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Variation in growth is the variance of student growth across classrooms, excluding variation due to measurement error. Of the full analysis sample of 657 classrooms in grades 1 to 3, reading comprehension was measured in grades 2 and 3 (66 percent of the sample).

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

C. Relationships that account for prerequisite actions

Certain types of instructional practices could occur only if teachers performed a specific prerequisite action. For example, teachers could engage students in defining new words during reading only if reading instruction, the prerequisite action, occurred. Teachers could have a low score on this instructional practice because they did not focus on vocabulary during reading or they did not do much reading instruction at all during the observation sessions. In general, for practices that depend on prerequisite actions, the relationships we presented in Chapter III of the main report treated teachers who did not perform the prerequisite action the same as teachers who performed the prerequisite action but did not perform the practice. Researchers may, however, be interested in relationships between instructional practices and student growth only in instances in which prerequisite actions occur. Following the previous example, researchers may want to answer the question, "When reading occurs, is engaging students in defining new words during reading associated with student growth?" This type of question asks whether

student growth depends on *how*, rather than *how frequently*, a teacher carries out reading instruction.

In this section, we present relationships between instructional practices and student growth when prerequisite actions occurred. To estimate these relationships, we included the frequency of the prerequisite actions as additional covariates in the regressions for measuring relationships between practices and student growth.

Nine of the 13 instructional practices examined by this study depended on prerequisite actions. Two practices—focusing on phonics and grammar during reading and helping students make connections between their prior knowledge and texts—could occur only if any reading-related instruction (pre-reading, reading, or post-reading) occurred. Three practices that engaged students in defining new words could occur only at the three specific phases of a reading lesson—pre-reading, reading, or post-reading; a fourth vocabulary-related practice, engaging students in defining new words outside of reading, could only occur during non-reading instruction. Likewise, three practices that focused on the meaning of texts could occur only at the three specific phases of a reading lesson. Being dependent on a prerequisite action limited the frequency with which a practice could occur. Reading-related instruction occurred in no more than half of the observation segments, although at least three-quarters of the segments included non-reading instruction (Table D.10).

Table D.10. Frequency of prerequisite actions

	Percentage of observation segments in which action was observed	
Prerequisite action	**Prekindergarten and kindergarten**	**Grades 1 through 3**
Any reading-related instruction	32	49
Pre-reading	25	36
Reading	30	46
Post-reading	22	32
Any non-reading instruction	85	76
Number of classrooms	**378**	**657**

Source: Authors' calculations from classroom observation data.

Relationships between practices and student growth when accounting for prerequisite actions were similar to those reported in Chapter III that did not account for prerequisite actions (Tables D.11 through D.14). Of the eight relationships that were positive and significant in the main analysis, five remained so in the analysis accounting for prerequisite actions. Three practices that depended on reading instruction—engaging students in defining new words during reading, focusing on the meaning of texts during reading, and helping students make connections between their prior knowledge and texts—no longer had significant relationships with background knowledge growth in the upper grades when comparing teachers who had the same frequency of reading instruction. However, the first two of those practices already had negative relationships with other outcomes and, therefore, could not have been identified as potentially promising; the third practice already had another positive and significant relationship with a different outcome and, therefore, remained potentially promising.

Moreover, one additional relationship was positive and significant when the analysis accounted for prerequisite actions. Focusing on the meaning of texts during pre-reading was associated with listening comprehension growth in the lower grades when comparing teachers with the same observed frequency of pre-reading instruction. This practice already had one other positive and significant relationship in this grade span (in both the main results and when accounting for prerequisite actions), so it appeared even more promising when the analysis accounted for prerequisite actions.

In summary, although the significance of some relationships changed when accounting for prerequisite actions, the practices identified as potentially promising did not change. These results are summarized in the third column in each of Tables D.15 and D.16.

Table D.11. Relationships between instructional practices that focus on phonics and grammar during reading and student growth in language and comprehension when accounting for prerequisite actions

	Prekindergarten and kindergarten		Grades 1 to 3	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value
When any reading-related instruction occurred, association between focusing on phonics and grammar during reading and				
Basic language skills	0.00	0.88	0.02	0.51
Background knowledge	-0.01	0.88	-0.04	0.59
Listening comprehension	-0.03	0.49	0.04	0.33
Reading comprehension			0.02	0.76
Number of classrooms	**378**		**220–657**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.12. Relationships between instructional practices that engage students in defining new words and student growth in language and comprehension when accounting for prerequisite actions

Student outcome	Prekindergarten and kindergarten		Grades 1 to 3	
	Relationship size	*p*-value	Relationship size	*p*-value
When pre-reading occurred, association between engaging students in defining new words during pre-reading and				
Basic language skills	-0.05*	0.08	-0.05	0.11
Background knowledge	-0.11***	<0.01	-0.08*	0.10
Listening comprehension	-0.04	0.17	0.03	0.29
Reading comprehension			0.01	0.60
When reading occurred, association between engaging students in defining new words during reading and				
Basic language skills	0.06*	0.06	-0.04*	0.09
Background knowledge	-0.04	0.37	0.05	0.12
Listening comprehension	0.04	0.36	0.02	0.64
Reading comprehension			-0.06***	<0.01
When post-reading occurred, association between engaging students in defining new words during post-reading and[a]				
Basic language skills			0.02	0.66
Background knowledge			0.01	0.85
Listening comprehension			0.06	0.19
Reading comprehension			0.15***	<0.01
When nonreading instruction occurred, association between engaging students in defining new words outside of reading and				
Basic language skills	-0.05	0.17	0.03	0.13
Background knowledge	0.00	0.98	0.06	0.14
Listening comprehension	0.02	0.60	-0.01	0.71
Reading comprehension			-0.01	0.59
Number of classrooms	**378**		**220–657**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

[a]The summary measure of engaging students in defining new words during post-reading did not vary across prekindergarten and kindergarten classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.13. Relationships between instructional practices that focus on the meaning of texts and student growth in language and comprehension when accounting for prerequisite actions

	Prekindergarten and kindergarten		Grades 1 to 3	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value
When pre-reading occurred, association between focusing on the meaning of texts during pre-reading and				
Basic language skills	0.09**	0.01	-0.02	0.26
Background knowledge	0.02	0.60	-0.09	0.11
Listening comprehension	0.08**	0.03	0.01	0.76
Reading comprehension			-0.03	0.31
When reading occurred, association between focusing on the meaning of texts during reading and				
Basic language skills	0.03	0.34	-0.06**	0.01
Background knowledge	-0.10*	0.05	0.03	0.45
Listening comprehension	0.05	0.22	-0.01	0.69
Reading comprehension			-0.04	0.39
When post-reading occurred, association between focusing on the meaning of texts during post-reading and				
Basic language skills	0.03	0.36	-0.01	0.60
Background knowledge	0.00	0.98	0.00	0.93
Listening comprehension	0.05	0.13	0.01	0.79
Reading comprehension			-0.03	0.47
Number of classrooms	**378**		**220–657**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.14. Relationships between instructional practices that help students use comprehension strategies and student growth in language and comprehension when accounting for prerequisite actions

	Prekindergarten and kindergarten		Grades 1 to 3	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value
When any reading-related instruction occurred, association between helping students make connections between their prior knowledge and texts and				
Basic language skills	0.03	0.34	0.03	0.16
Background knowledge	0.04	0.27	0.04	0.36
Listening comprehension	0.11***	<0.01	0.03	0.42
Reading comprehension			0.06*	0.06
Number of classrooms	**378**		**220–657**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

D. Summary of potentially promising practices across alternative analyses

Among the potentially promising practices identified in the main analyses, we found that certain practices remained potentially promising in many alternative analyses, whereas others did not. Tables D.15 and D.16 present a summary of the potentially promising practices identified by the main analysis and each of the alternative analyses discussed in this appendix, including analyses that:

- Accounted for the other practices that teachers used when examining relationships between each practice and student growth (see detailed results in Tables D.6 through D.9)
- Accounted for prerequisite actions (see detailed results in Tables D.11 through D.14)
- Required positive relationships to be statistically significant at the 5 percent level (see detailed results in Tables D.1 through D.5)
- Required positive relationships to be statistically significant at the 5 percent level after taking into account the number of relationships examined for each instructional practice (see detailed results in Tables D.1 through D.5)

Of the five potentially promising practices identified in the main analyses for prekindergarten and kindergarten, one remained potentially promising in all relevant alternative analyses, two remained potentially promising in at least half (but not all) of the alternative analyses, and two remained potentially promising in fewer than half of the alternative analyses.

Of the five potentially promising practices identified in the main analyses for grades 1 to 3, two remained potentially promising in all relevant alternative analyses, two remained potentially promising in at least half (but not all) of the alternative analyses, and one remained potentially promising in fewer than half of the alternative analyses. Chapter III, Section B provides a more detailed discussion of these tiers of potentially promising practices.

Table D.15. Instructional practices identified as potentially promising in prekindergarten and kindergarten under alternative analyses

Instructional practice	Promising practice when:				
	Using main approach?	Accounting for other practices?	Accounting for prerequisite actions?	Requiring positive relationships to be significant at the 5 percent level?	Adjusting the significance of positive relationships for the number of relationships examined?
1. Encouraging students' oral language	No	No	NA[b]	No	No
2. Focusing on phonics and grammar during reading	No	No	No	No	No
3. Engaging students in defining new words during pre-reading	No	No	No	No	No
4. Engaging students in defining new words during reading	Yes	No	Yes	Yes	No
5. Engaging students in defining new words during post-reading	NA[a]	NA[a]	NA[a]	NA[a]	NA[a]
6. Engaging students in defining new words outside of reading	No	No	No	No	No
7. Focusing on the meaning of texts during pre-reading	Yes	Yes	Yes	Yes	Yes
8. Focusing on the meaning of texts during reading	No	No	No	No	No
9. Focusing on the meaning of texts during post-reading	No	No	No	No	No
10. Helping students make connections between their prior knowledge and texts	Yes	Yes	Yes	Yes	No
11. Teaching students to use other comprehension strategies	No	No	NA[b]	No	No
12. Focusing on world knowledge	Yes	No	NA[b]	No	No
13. Focusing on higher-order thinking	Yes	No	NA[b]	Yes	No

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The table includes data from 378 prekindergarten and kindergarten classrooms.

A practice is considered potentially promising if there was at least one positive and significant relationship and no negative and significant relationships.

[a] Within the lower grades, we did not find evidence that teachers in the study differed in the usual extent to which they engaged students in defining new words during post-reading. Therefore, we did not examine the relationship between this practice and student growth in the lower grades.

[b] Some practices do not have prerequisite actions and therefore have no results for this alternative analysis.

NA = not applicable.

Table D.16. Instructional practices identified as potentially promising in grades 1 to 3 under alternative analyses

Instructional practice	Promising practice when:				
	Using main approach?	Accounting for other practices?	Accounting for prerequisite actions?	Requiring positive relationships to be significant at the 5 percent level?	Adjusting the significance of positive relationships for the number of relationships examined?
1. Encouraging students' oral language	Yes	No	NA[a]	Yes	No
2. Focusing on phonics and grammar during reading	No	No	No	No	No
3. Engaging students in defining new words during pre-reading	No	No	No	No	No
4. Engaging students in defining new words during reading	No	No	No	No	No
5. Engaging students in defining new words during post-reading	Yes	Yes	Yes	Yes	Yes
6. Engaging students in defining new words outside of reading	No	No	No	No	No
7. Focusing on the meaning of texts during pre-reading	No	No	No	No	No
8. Focusing on the meaning of texts during reading	No	No	No	No	No
9. Focusing on the meaning of texts during post-reading	No	No	No	No	No
10. Helping students make connections between their prior knowledge and texts	Yes	Yes	Yes	Yes	No
11. Teaching students to use other comprehension strategies	Yes	Yes	NA[a]	Yes	No
12. Focusing on world knowledge	No	No	NA[a]	No	No
13. Focusing on higher-order thinking	Yes	Yes	NA[a]	Yes	Yes

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: Of the full analysis sample of 657 classrooms in grades 1 to 3, background knowledge was measured in grade 1 (33 percent of the sample), reading comprehension was measured in grades 2 and 3 (66 percent of the sample), and the remaining two outcomes were measured in all classrooms.

A practice is considered potentially promising if there was at least one positive and significant relationship and no negative and significant relationships.

[a] Some practices do not have prerequisite actions and therefore have no results for this alternative analysis.

NA = not applicable.

E. Detailed findings on relationships within student subgroups

In Chapter III of the main report, we summarized findings for the relationships between instructional practices and student growth within subgroups defined by students' home language (English or non-English) and baseline test score (top 40 percent or bottom 40 percent). As discussed in the main report, potentially promising practices differed across subgroups. Tables D.17 through D.31 present detailed findings, including the size and statistical significance of the relationships for each subgroup. In addition to providing results by home language (Tables D.17 through D.21) and baseline test score (Tables D.27 through D.31), these tables also provide results by gender (Tables D.22 through D.26).

Table D.17. Relationships between instructional practices that encourage students' oral language or focus on phonics and grammar during reading and student growth in language and comprehension, by home language

	Prekindergarten and kindergarten				Grades 1 to 3			
	English home language		Non-English home language		English home language		Non-English home language	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between encouraging students' oral language and								
Basic language skills	0.00	0.92	-0.02	0.63	0.01	0.71	0.04	0.25
Background knowledge	0.00	0.93	0.04	0.45	0.09**	0.01	0.06	0.64
Listening comprehension	-0.04	0.30	0.07	0.12	-0.01	0.77	0.03	0.55
Reading comprehension					0.00	0.96	-0.03	0.58
Association between focusing on phonics and grammar during reading and								
Basic language skills	-0.02	0.61	0.00	0.96	-0.01	0.61	0.04	0.40
Background knowledge	-0.03	0.56	-0.05	0.53	0.09	0.14	-0.03	0.85
Listening comprehension	-0.10**	0.02	0.00	0.99	-0.01	0.80	0.11**	0.04
Reading comprehension					0.00	1.00	0.07	0.32
Number of classrooms	**346–348**		**203–214**		**199–607**		**112–348**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.18. Relationships between instructional practices that engage students in defining new words and student growth in language and comprehension, by home language

	Prekindergarten and kindergarten				Grades 1 to 3			
	English home language		Non-English home language		English home language		Non-English home language	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between engaging students in defining new words during pre-reading and								
Basic language skills	-0.07***	<0.01	0.06*	0.08	-0.02	0.37	-0.10	0.11
Background knowledge	-0.12***	<0.01	-0.09	0.20	0.03	0.55	-0.09	0.39
Listening comprehension	-0.06*	0.05	-0.03	0.50	0.02	0.65	0.07	0.22
Reading comprehension					0.00	0.91	0.06	0.40
Association between engaging students in defining new words during reading and								
Basic language skills	0.03	0.25	0.08	0.16	-0.02	0.53	-0.05**	0.05
Background knowledge	-0.05	0.29	-0.05	0.39	0.12***	<0.01	-0.01	0.94
Listening comprehension	-0.04	0.34	0.11**	0.04	0.03	0.44	0.03	0.40
Reading comprehension					-0.06	0.11	0.00	0.99
Association between engaging students in defining new words during post-reading and[a]								
Basic language skills					0.01	0.90	0.02	0.75
Background knowledge					0.09	0.11	0.14	0.44
Listening comprehension					0.02	0.58	0.12**	0.05
Reading comprehension					0.04	0.68	0.21***	<0.01
Association between engaging students in defining new words outside of reading and								
Basic language skills	-0.06	0.18	-0.02	0.73	0.06*	0.07	0.01	0.75
Background knowledge	-0.02	0.80	0.01	0.93	0.05	0.33	-0.08	0.60
Listening comprehension	0.02	0.62	-0.01	0.84	0.02	0.37	-0.02	0.66
Reading comprehension					-0.04	0.27	0.02	0.59
Number of classrooms	**346–348**		**203–214**		**199–607**		**112–348**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

[a] The summary measure of engaging students in defining new words during post-reading did not vary across prekindergarten and kindergarten classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.19. Relationships between instructional practices that focus on the meaning of texts and student growth in language and comprehension, by home language

	Prekindergarten and kindergarten				Grades 1 to 3			
	English home language		Non-English home language		English home language		Non-English home language	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between focusing on the meaning of texts during pre-reading and								
Basic language skills	0.07*	0.05	0.11*	0.08	-0.04	0.18	0.00	0.94
Background knowledge	-0.02	0.80	0.09	0.19	-0.03	0.52	-0.05	0.68
Listening comprehension	-0.03	0.54	0.25***	<0.01	-0.01	0.69	0.06	0.15
Reading comprehension					-0.06	0.14	0.01	0.81
Association between focusing on the meaning of texts during reading and								
Basic language skills	-0.02	0.63	0.07	0.19	-0.08***	<0.01	0.03	0.35
Background knowledge	-0.11**	0.02	-0.04	0.62	0.07	0.10	0.02	0.86
Listening comprehension	-0.08*	0.07	0.13**	0.03	-0.03	0.37	0.10**	0.02
Reading comprehension					-0.04	0.30	0.05	0.36
Association between focusing on the meaning of texts during post-reading and								
Basic language skills	-0.01	0.76	0.07	0.25	-0.05	0.11	0.05	0.18
Background knowledge	-0.06	0.25	0.10	0.19	0.12**	0.01	0.04	0.67
Listening comprehension	-0.08**	0.04	0.21***	<0.01	-0.04	0.20	0.12**	0.03
Reading comprehension					-0.05	0.21	0.04	0.52
Number of classrooms	**346–348**		**203–214**		**199–607**		**112–348**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.20. Relationships between instructional practices that help students use comprehension strategies and student growth in language and comprehension, by home language

	Prekindergarten and kindergarten				Grades 1 to 3			
	English home language		Non-English home language		English home language		Non-English home language	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between helping students make connections between their prior knowledge and texts and								
Basic language skills	0.07***	<0.01	0.02	0.81	0.01	0.72	0.07	0.13
Background knowledge	0.07	0.26	0.06	0.40	0.08	0.15	0.16	0.19
Listening comprehension	0.03	0.44	0.20***	<0.01	-0.05	0.13	0.13***	<0.01
Reading comprehension					0.03	0.44	0.13*	0.05
Association between teaching students to use other comprehension strategies and								
Basic language skills	0.01	0.73	-0.07	0.25	0.02	0.43	-0.01	0.70
Background knowledge	-0.04	0.53	-0.02	0.81	0.06	0.16	-0.46**	0.02
Listening comprehension	0.01	0.78	-0.06	0.44	0.04**	0.04	0.09*	0.08
Reading comprehension					0.01	0.82	0.06	0.16
Number of classrooms	**346–348**		**203–214**		**199–607**		**112–348**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.21. Relationships between instructional practices that focus on world knowledge or higher-order thinking and student growth in language and comprehension, by home language

	Prekindergarten and kindergarten				Grades 1 to 3			
	English home language		Non-English home language		English home language		Non-English home language	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between focusing on world knowledge and								
Basic language skills	0.03	0.15	0.09*	0.06	-0.04	0.11	-0.06**	0.03
Background knowledge	0.00	0.93	0.08**	0.04	0.07	0.13	0.04	0.72
Listening comprehension	-0.02	0.54	0.15**	0.04	-0.01	0.87	-0.06	0.13
Reading comprehension					-0.02	0.49	-0.01	0.78
Association between focusing on higher-order thinking and								
Basic language skills	0.05**	0.01	0.04	0.26	0.02	0.52	-0.02	0.65
Background knowledge	0.07	0.11	0.06	0.34	0.10**	0.02	0.07	0.36
Listening comprehension	0.04	0.29	0.11	0.17	0.02	0.46	-0.12**	0.03
Reading comprehension					0.00	0.96	-0.01	0.82
Number of classrooms	**346–348**		**203–214**		**199–607**		**112–348**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.22. Relationships between instructional practices that encourage students' oral language or focus on phonics and grammar during reading and student growth in language and comprehension, by gender

	Prekindergarten and kindergarten				Grades 1 to 3			
	Boys		Girls		Boys		Girls	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between encouraging students' oral language and								
Basic language skills	-0.07***	<0.01	0.08***	<0.01	0.01	0.77	0.01	0.64
Background knowledge	-0.05	0.24	0.03	0.44	0.10*	0.07	0.09*	0.08
Listening comprehension	-0.05	0.22	0.02	0.70	0.04	0.21	-0.03	0.39
Reading comprehension					0.00	0.91	-0.01	0.89
Association between focusing on phonics and grammar during reading and								
Basic language skills	0.02	0.69	0.00	0.98	0.03	0.45	-0.01	0.65
Background knowledge	-0.05	0.32	-0.03	0.57	0.05	0.40	0.08	0.42
Listening comprehension	-0.13***	<0.01	-0.03	0.50	0.02	0.61	0.06	0.35
Reading comprehension					0.03	0.37	0.00	0.97
Number of classrooms	**345–350**		**334–339**		**200–603**		**207–617**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.23. Relationships between instructional practices that engage students in defining new words and student growth in language and comprehension, by gender

	Prekindergarten and kindergarten				Grades 1 to 3			
	Boys		Girls		Boys		Girls	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between engaging students in defining new words during pre-reading and								
Basic language skills	-0.04*	0.08	-0.03	0.52	0.02	0.54	-0.11***	<0.01
Background knowledge	-0.08**	0.02	-0.15***	<0.01	0.09	0.20	-0.12**	0.03
Listening comprehension	-0.08**	0.03	-0.06	0.14	0.07*	0.06	0.02	0.56
Reading comprehension					0.03	0.55	0.00	0.99
Association between engaging students in defining new words during reading and								
Basic language skills	0.07**	0.02	0.06	0.25	-0.01	0.77	-0.07**	0.02
Background knowledge	-0.09*	0.06	0.00	0.97	0.04	0.33	0.15**	0.01
Listening comprehension	-0.04	0.49	0.06	0.28	0.02	0.54	0.02	0.68
Reading comprehension					0.01	0.74	-0.08**	0.04
Association between engaging students in defining new words during post-reading and[a]								
Basic language skills					0.04	0.44	0.00	0.93
Background knowledge					0.09*	0.10	0.01	0.94
Listening comprehension					0.06	0.24	0.10	0.14
Reading comprehension					0.15**	0.03	0.11	0.12
Association between engaging students in defining new words outside of reading and								
Basic language skills	-0.04	0.41	-0.08*	0.09	0.03	0.35	0.02	0.46
Background knowledge	-0.03	0.66	0.02	0.70	0.07	0.23	0.05	0.47
Listening comprehension	0.03	0.62	-0.02	0.70	0.02	0.64	-0.04	0.21
Reading comprehension					-0.06*	0.09	0.01	0.85
Number of classrooms	**345–350**		**334–339**		**200–603**		**207–617**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

[a] The summary measure of engaging students in defining new words during post-reading did not vary across prekindergarten and kindergarten classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.24. Relationships between instructional practices that focus on the meaning of texts and student growth in language and comprehension, by gender

	Prekindergarten and kindergarten				Grades 1 to 3			
	Boys		Girls		Boys		Girls	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between focusing on the meaning of texts during pre-reading and								
Basic language skills	0.10*	0.07	0.09**	0.03	0.00	0.86	-0.02	0.26
Background knowledge	0.04	0.51	-0.09	0.18	-0.08	0.36	-0.04	0.57
Listening comprehension	0.05	0.40	0.02	0.67	-0.02	0.53	0.04	0.39
Reading comprehension					0.02	0.68	-0.07	0.11
Association between focusing on the meaning of texts during reading and								
Basic language skills	0.03	0.47	0.01	0.73	-0.06**	0.02	-0.04	0.24
Background knowledge	-0.13**	0.01	-0.08	0.12	-0.02	0.61	0.17**	0.02
Listening comprehension	-0.07	0.16	0.00	0.95	-0.03	0.40	0.01	0.81
Reading comprehension					-0.01	0.85	-0.01	0.70
Association between focusing on the meaning of texts during post-reading and								
Basic language skills	0.01	0.86	0.05	0.38	0.01	0.87	-0.01	0.72
Background knowledge	-0.01	0.77	-0.01	0.83	0.02	0.79	0.13**	0.04
Listening comprehension	-0.01	0.80	0.02	0.66	-0.03	0.36	0.06	0.16
Reading comprehension					0.01	0.86	-0.03	0.67
Number of classrooms	**345–350**		**334–339**		**200–603**		**207–617**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.25. Relationships between instructional practices that help students use comprehension strategies and student growth in language and comprehension, by gender

	Prekindergarten and kindergarten				Grades 1 to 3			
	Boys		Girls		Boys		Girls	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between helping students make connections between their prior knowledge and texts and								
Basic language skills	0.07	0.15	0.04	0.37	0.07***	<0.01	-0.01	0.84
Background knowledge	0.15**	0.01	-0.10*	0.06	0.04	0.56	0.11*	0.06
Listening comprehension	0.11*	0.07	0.06	0.11	0.00	0.91	0.05	0.30
Reading comprehension					0.07	0.22	0.06	0.19
Association between teaching students to use other comprehension strategies and								
Basic language skills	0.05	0.18	-0.06	0.30	0.07**	0.02	-0.03	0.29
Background knowledge	0.02	0.68	-0.07	0.31	0.03	0.60	-0.01	0.87
Listening comprehension	-0.05	0.43	-0.02	0.73	0.01	0.62	0.08**	0.04
Reading comprehension					0.04	0.34	0.01	0.69
Number of classrooms	**345–350**		**334–339**		**200–603**		**207–617**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.26. Relationships between instructional practices that focus on world knowledge or higher-order thinking and student growth in language and comprehension, by gender

	Prekindergarten and kindergarten				Grades 1 to 3			
	Boys		Girls		Boys		Girls	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between focusing on world knowledge and								
Basic language skills	0.01	0.62	0.09**	0.03	-0.07**	0.02	-0.04*	0.06
Background knowledge	-0.05	0.25	0.04	0.43	0.02	0.66	0.09	0.19
Listening comprehension	0.00	0.94	0.06	0.18	-0.07**	0.04	0.04	0.30
Reading comprehension					0.01	0.85	-0.04	0.17
Association between focusing on higher-order thinking and								
Basic language skills	0.03	0.31	0.05**	0.04	-0.03	0.32	0.02	0.53
Background knowledge	0.00	0.96	0.07**	0.04	0.10**	0.02	0.13**	0.04
Listening comprehension	-0.01	0.88	0.10***	<0.01	-0.05	0.18	0.00	0.93
Reading comprehension					-0.01	0.69	0.01	0.78
Number of classrooms	**345–350**		**334–339**		**200–603**		**207–617**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.27. Relationships between instructional practices that encourage students' oral language or focus on phonics and grammar during reading and student growth in language and comprehension, by baseline test score

	Prekindergarten and kindergarten				Grades 1 to 3			
	High achievers		Low achievers		High achievers		Low achievers	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between encouraging students' oral language and								
Basic language skills	0.00	1.00	-0.01	0.79	-0.05*	0.07	0.06**	0.01
Background knowledge	-0.09	0.29	-0.01	0.74	-0.03	0.57	0.16**	0.03
Listening comprehension	0.00	0.94	-0.07	0.25	-0.01	0.65	0.00	0.96
Reading comprehension					-0.09***	<0.01	0.05	0.11
Association between focusing on phonics and grammar during reading and								
Basic language skills	-0.04	0.21	0.02	0.49	-0.02	0.60	0.02	0.68
Background knowledge	-0.03	0.72	-0.06	0.23	-0.04	0.70	0.11	0.26
Listening comprehension	-0.02	0.70	-0.23***	<0.01	0.03	0.48	-0.02	0.79
Reading comprehension					-0.03	0.68	0.06	0.11
Number of classrooms	**292–325**		**286–321**		**163–582**		**184–532**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. High and low achievers are those whose fall test scores were in, respectively, the top and bottom 40 percent of students in the study.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.28. Relationships between instructional practices that engage students in defining new words and student growth in language and comprehension, by baseline test score

	Prekindergarten and kindergarten				Grades 1 to 3			
	High achievers		Low achievers		High achievers		Low achievers	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between engaging students in defining new words during pre-reading and								
Basic language skills	-0.07***	<0.01	0.04	0.39	-0.09**	0.01	0.02	0.77
Background knowledge	-0.13	0.13	-0.19**	0.05	-0.01	0.85	-0.02	0.76
Listening comprehension	-0.08***	<0.01	-0.09	0.12	-0.02	0.59	0.02	0.66
Reading comprehension					-0.02	0.76	0.02	0.65
Association between engaging students in defining new words during reading and								
Basic language skills	0.07**	0.02	0.07	0.20	-0.09***	<0.01	0.04	0.27
Background knowledge	-0.14**	0.02	-0.03	0.58	0.05	0.41	0.15**	0.03
Listening comprehension	0.01	0.93	0.05	0.41	-0.06	0.18	0.02	0.61
Reading comprehension					-0.04	0.36	-0.02	0.56
Association between engaging students in defining new words during post-reading and[a]								
Basic language skills					0.02	0.72	0.09	0.35
Background knowledge					-0.03	0.71	0.26*	0.06
Listening comprehension					0.03	0.62	0.05	0.46
Reading comprehension					-0.01	0.89	0.22***	<0.01
Association between engaging students in defining new words outside of reading and								
Basic language skills	-0.04	0.30	-0.06	0.25	-0.02	0.48	0.10***	<0.01
Background knowledge	0.02	0.81	-0.05	0.45	-0.01	0.87	0.10	0.21
Listening comprehension	0.03	0.60	0.03	0.76	-0.06*	0.06	0.01	0.70
Reading comprehension					-0.02	0.57	0.00	0.97
Number of classrooms	**292–325**		**286–321**		**163–582**		**184–532**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. High and low achievers are those whose fall test scores were in, respectively, the top and bottom 40 percent of students in the study.

[a] The summary measure of engaging students in defining new words during post-reading did not vary across prekindergarten and kindergarten classrooms.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.29. Relationships between instructional practices that focus on the meaning of texts and student growth in language and comprehension, by baseline test score

	Prekindergarten and kindergarten				Grades 1 to 3			
	High achievers		Low achievers		High achievers		Low achievers	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between focusing on the meaning of texts during pre-reading and								
Basic language skills	0.14***	<0.01	0.11*	0.05	0.00	0.90	-0.01	0.82
Background knowledge	-0.03	0.78	-0.02	0.79	-0.16**	0.02	0.04	0.78
Listening comprehension	0.03	0.64	0.00	0.98	-0.04	0.25	0.06*	0.09
Reading comprehension					-0.05	0.48	-0.01	0.87
Association between focusing on the meaning of texts during reading and								
Basic language skills	-0.01	0.85	0.08*	0.07	-0.02	0.70	-0.01	0.70
Background knowledge	-0.15*	0.06	-0.12*	0.08	0.03	0.62	0.13**	0.03
Listening comprehension	-0.04	0.54	0.02	0.84	0.00	0.95	-0.04	0.37
Reading comprehension					-0.01	0.88	-0.03	0.44
Association between focusing on the meaning of texts during post-reading and								
Basic language skills	0.01	0.89	0.06	0.11	0.04	0.35	-0.03	0.51
Background knowledge	-0.01	0.95	-0.03	0.43	0.01	0.88	0.07	0.42
Listening comprehension	0.00	0.99	-0.03	0.75	-0.01	0.81	0.00	0.97
Reading comprehension					0.00	0.95	-0.02	0.74
Number of classrooms	**292–325**		**286–321**		**163–582**		**184–532**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. High and low achievers are those whose fall test scores were in, respectively, the top and bottom 40 percent of students in the study.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.30. Relationships between instructional practices that help students use comprehension strategies and student growth in language and comprehension, by baseline test score

	Prekindergarten and kindergarten				Grades 1 to 3			
	High achievers		Low achievers		High achievers		Low achievers	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between helping students make connections between their prior knowledge and texts and								
Basic language skills	0.10**	0.01	-0.01	0.93	0.05	0.22	0.07	0.21
Background knowledge	0.06	0.54	0.01	0.82	0.01	0.85	0.16	0.14
Listening comprehension	0.11**	0.04	0.02	0.69	-0.05	0.26	0.01	0.88
Reading comprehension					0.10	0.12	0.05	0.24
Association between teaching students to use other comprehension strategies and								
Basic language skills	0.00	0.94	-0.04	0.44	-0.01	0.76	0.04	0.17
Background knowledge	0.00	0.97	0.04	0.63	-0.06	0.34	0.01	0.94
Listening comprehension	0.04	0.53	-0.08	0.47	0.03	0.34	0.07**	0.02
Reading comprehension					-0.01	0.89	0.04	0.33
Number of classrooms	**292–325**		**286–321**		**163–582**		**184–532**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. High and low achievers are those whose fall test scores were in, respectively, the top and bottom 40 percent of students in the study.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

Table D.31. Relationships between instructional practices that focus on world knowledge or higher-order thinking and student growth in language and comprehension, by baseline test score

	Prekindergarten and kindergarten				Grades 1 to 3			
	High achievers		Low achievers		High achievers		Low achievers	
Student outcome	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value	Relationship size	*p*-value
Association between focusing on world knowledge and								
Basic language skills	0.07***	<0.01	0.05	0.35	-0.08**	0.01	-0.04	0.30
Background knowledge	-0.12	0.16	-0.03	0.52	0.09	0.23	0.06	0.45
Listening comprehension	0.07*	0.07	-0.05	0.29	-0.06	0.14	-0.03	0.45
Reading comprehension					0.03	0.49	-0.03	0.43
Association between focusing on higher-order thinking and								
Basic language skills	0.06**	0.05	0.03	0.49	-0.03*	0.07	0.01	0.86
Background knowledge	0.02	0.84	0.02	0.41	0.08	0.21	0.11	0.14
Listening comprehension	0.04	0.51	-0.02	0.75	-0.05**	0.05	0.01	0.82
Reading comprehension					-0.01	0.82	0.04	0.34
Number of classrooms	**292–325**		**286–321**		**163–582**		**184–532**	

Source: Authors' calculations using data from the fall and spring tests administered by the study team and classroom observations conducted by the study team.

Note: The relationship size is the change in student test scores, measured in student-level standard deviations, that is associated with a one standard deviation increase in the instructional practice across classrooms. High and low achievers are those whose fall test scores were in, respectively, the top and bottom 40 percent of students in the study.

*Significantly different from zero at the .10 level, two-tailed test.

**Significantly different from zero at the .05 level, two-tailed test.

***Significantly different from zero at the .01 level, two-tailed test.

APPENDIX E

CLASSROOM OBSERVATION RUBRIC AND CODING FORM

This page has been left blank for double-sided copying.

OBSERVATION OF LANGUAGE AND LITERACY INSTRUCTION (OLLI)

Early Childhood Language Development

RUBRIC AND CODING FORM

School Name: _____

Teacher Name: _____

Bar Code:

Observation Form: 1 ☐ 2 ☐ 3 ☐ 4 ☐

Observer Name: _____

Date of Observation: |__|__| / |__|__| / |_2_|_0_|_1_|_2_|
Month Day Year

OVERALL PROCEDURES

1	***Identify the lead teacher.***	*See Guidelines on page 2.*
2	***Conduct a 1-minute Classroom Scan (SCAN).***	*Remember to record the start time, number of adults and children in the classroom and all the structures and activities of all children in the classroom during the 1 minute.*
3	***Set your timer on 15-minutes and begin your observation.***	*Take notes in your note-taking booklet. Focus on the lead teacher (see Guidelines on page 2). Focus on all activities the lead teacher is involved in.*
4	***When the 15-minutes is up, code for DESC.***	*Record the end time, the number of children and adults you focused on during that segment, and make note of any unusual circumstances that occurred during the segment.*
5	***Complete the codes for dimensions LANG through HIGH.***	*Only code what you see. Aim to complete your coding in 10-15 minutes per observation segment.*
6	***Begin another observation segment, followed by coding.***	*SCAN for 1 minute, observe/take notes for 15 minutes, then code for DESC through HIGH.*
7	***Continue the cycle until you have completed 6 observation segments.***	*This should take about 3 hours.*
8	***After you have observed six segments, complete codes for SUMM.***	*Base your SUMM codes on your observation across all six segments.*

OLLI KEY WORDS AND PHRASES

Adult-led Instruction	*This is when the teacher is leading one or more students in a learning activity. This does not include times when the teacher is monitoring students working independently or in groups.*
Adult/student Interaction	*Adults can interact with students by talking with them and/or monitoring their work. If an adult walks around the room watching the students that is considered interaction.*
Adult vs. teacher	*The terms adult and teacher are used interchangeably in this rubric.*

GUIDELINES FOR OVERALL PROCEDURES

1	***Who should I focus on during the observation?***	*Generally speaking, we will focus on one adult throughout the observation period (all six segments); this will typically be the lead teacher. The following table helps you identify the lead teacher and provides guidance on who to follow in special situations.*
	IDENTIFYING THE LEAD TEACHER	
2	***If your observation assignment doesn't indicate who the lead teacher is.***	*Arrive early so that you can ask your field supervisor.*
3	***If your field supervisor isn't available or doesn't know who the lead teacher is.***	*Arrive early so that you can ask the adults in the classroom, before instruction begins.*
4	***You arrive in the middle of an activity or lesson and you do not know who the lead teacher is.***	*Assume that the adult leading the instructional activity is the lead teacher and follow him or her for the rest of the observation segment. Then, try to confirm who the lead teacher is for the next segment and follow the lead teacher.*
5	***You arrive in the middle of an activity/lesson, you don't know who the lead teacher is, and more than one adult is leading different groups in activities.***	*Assume that the adult leading the instructional activity with the most students is the lead teacher and follow him or her for the rest of the observation segment. Then, try to confirm who the lead teacher is for the next segment and follow the lead teacher from that point on.*
6	***There are two lead teachers in a classroom: both are delivering the instruction (co-teaching).***	*Treat them as if they were one teacher; include what both say and do in your coding. If they divide the class into groups, follow the teacher leading the instructional activity with the most students. Continue to follow this teacher for the whole observation period.*
7	***During your observation, another adult joins the lead teacher in leading the instructional activity.***	*Treat them as if they are one teacher, the lead teacher, and include what both do and say in your coding.*
	WHERE TO FOCUS DURING DIFFERENT INSTRUCTIONAL STRUCTURES	
8	***The lead teacher works with one student while another adult works with other students (or even leads instruction with the other students).***	*Stay with the lead teacher.*
9	***A specialist or an assistant teacher is leading whole group instruction, while the lead teacher is part of the group and chimes in occasionally (or not at all).***	*Watch the lead teacher and code for his or her interactions when participating in the activity. Even if the lead teacher does not participate you would still be coding for him or her (not the other adult who is leading the instruction).*
10	***The lead teacher is not interacting with students, such as when completing paper work and/or the students are working independently (and the teacher isn't monitoring them).***	*Make a note of the amount of time the lead teacher is not interacting with students and watch the students to code for their interactions.*
11	***There is no teacher-led instruction and multiple teachers are monitoring/playing with students.***	*Stay with the lead teacher.*
12	***During your observation segment the lead teacher leaves the room.***	*If it is clear he or she will not be returning, begin following the teacher who is working with the most children at the time the lead teacher leaves. Stay with this teacher for the rest of the observation segment. If it is not clear whether he or she will return make a note of the time the lead teacher was not interacting with students and watch the students for their interactions. If the lead teacher doesn't return by the start of the next observation segment, select a new teacher to follow (see guidance on selecting the teacher above).*
	WHAT TO DO WHEN STUDENTS LEAVE THE CLASS	
13	***A large group of students leave to work with a teacher in another classroom (could be for academic instruction or for music, art, etc).***	*Stay with the lead teacher and students in the classroom. Never follow students out of the classroom unless the lead teacher goes with them. If you do need to leave the classroom to follow the lead teacher, make a note of this in D4.*
14	***All of the students leave to work with another teacher and a new group of students comes to work with the lead teacher.***	*Stay with the lead teacher and conduct the observation, focusing on the teacher and the students the teacher interacts with during the segment. If all of the students leave the classroom and a new group comes in, make a note of this in D4.*
15	***All of the students leave the classroom to work with another teacher (there are no students in the classroom).***	*If you have observed for at least 5 minutes (not including SCAN) before the students leave the classroom, stop the observation and code the segment as is. If you observed the class for less than 5 minutes before they leave do not code this observation segment. Either way, wait for the students to return to the lead teacher before beginning your next observation segment.*
	SPECIAL CIRCUMSTANCES	
16	***The teacher sets up a video, audio, or other media for all students to watch (such as a movie, television show, video game).***	*We consider this as teacher-provided instruction. Code as you would for other teacher-directed activities. Even if the teacher was not interacting with students (he or she is grading papers) you would consider the media as teacher-provided instruction. Make a note of this in D4.*
17	***Many of the day's lessons are conducted outdoors by the teacher.***	*Stay with the lead teacher and conduct the observation in the outdoor setting. Make a note of this in D4.*
18	***The lead teacher is absent, it is early in the week, and he or she is likely to return.***	*Report this to your field supervisor, so that the observation can be rescheduled for later that week (or as soon as possible the following week).*
19	***The lead teacher is out for an indeterminate time and a long-term substitute is assigned to the classroom.***	*Continue with the observation. Consider the substitute as the lead teacher. Make a note of this in D4.*
20	***The teacher does not want you to conduct the observation.***	*Do not try to negotiate with the teacher. Contact your field supervisor immediately and he or she will address the issue.*
21	***The segment is cut short but you have observed for at least 5 minutes (not including the one minute scan).***	*Code the segment based on that shortened observation time. If you observed for less than 5 minutes, do not code that segment; instead, please start a new observation segment as soon as possible.* ***Do not code any observation segment that is less than five minutes long.***

CLASSROOM SCAN (SCAN): ITEMS S1-S5

START EACH SEGMENT WITH THE CLASSROOM SCAN (SCAN). IT WILL TAKE YOU ONE MINUTE TO COMPLETE THE SCAN ITEMS. FIRST, RECORD THE START TIME (S1), COUNT THE TOTAL NUMBER OF CHILDREN IN THE ROOM (S2), AND COUNT THE TOTAL NUMBER OF ADULTS IN THE ROOM (S3). CONTINUE SCANNING THE CLASSROOM FOR 1 MINUTE, COUNTING THE TOTAL NUMBER OF ADULTS IN THE ROOM WHO INTERACT WITH CHILDREN (S4) AND RECORDING ALL THE CLASSROOM STRUCTURE(S) (S5) AND ACTIVITIES TAKING PLACE (S6). BE SURE TO CAPTURE THE ACTIVITIES OF ALL STUDENTS PRESENT DURING THE ONE-MINUTE SCAN. AFTER YOU COMPLETE YOUR SCAN, SET YOUR CLOCK FOR 15 MINUTES AND CONTINUE OBSERVING FOR 15 MORE MINUTES, TAKING NOTES AS YOU OBSERVE.

KEY WORDS AND PHRASES		
S4	***Adults who interact with children***	*This includes the teacher and any other adult in the room who are working with, speaking with or otherwise interacting with the children.*
S5	***Whole Group/ Whole Class***	*This refers to times when all the students in the class (the whole class) are engaged in the same teacher-led activity.*
	Large Group	*This refers to times when more than half of the students in the class are engaged in the same teacher-led activity.*
	Small Group	*This refers to an intentional grouping of 3 or more students who are working on the same activity. In a class of 24 students, the teacher may divide the students into 8 teams of 3 students, 6 teams of 4 students, 4 teams of 6 students, or 2 teams of 12. Students may be working on their own or with the teacher. Don't count spontaneous groupings of children playing together in centers as small group, instead code that as centers.*
	Centers	*This classroom configuration is found in most pre-K and some K classrooms, where there are clearly established areas of the classroom where students go to engage in specific activities (different activities in each center), such as the playing with blocks area, classroom library/reading area, computer games area, the kitchen play area, the sand table area. You may also see some of these centers in grades 1-3 classrooms (such as computer area and class library area). Students may be working on their own or with the teacher.*
	Partners/Pairs	*This refers to times when the teacher has asked students to work in pairs OR when students work in groups of two on their own or with a teacher. Think-pair-share is considered partner work (even if the whole class is doing it).*
	Individual	*This refers to times when students work alone and/or the teacher works with students one-on-one.*
	Other	*Any other configuration that you see that doesn't fit into the categories above.*

FAQs		
S3	***Do I count myself as an adult in the classroom for S3?***	*No, do not include yourself in the counts of adults.*
S5	***Can there be more than one type of structure at a time?***	*Yes. Please code any and all types of classroom structures that you see during your observation. For example, if 3/4 the class is in one group and the rest are in a small group, you would code large and small group. HOWEVER, if students are in centers, just code #4 for centers and do not code the other types of classroom structures.*
	What if the teacher divides a class of 24 students evenly into two groups, does this count as large or small group?	*If the group is split evenly, you would code this as small groups. Even though a group of 12 students doesn't seem that small, working with 12 other students is a smaller group than working with 24 other students.* *If the groups were split so that there were 13 in one group and 11 in the other you would code large group and small group (because one group would have a more than half of the children in it).*
	When students are seated together at a table, is this always coded as small group?	*Only if they are working together. In some classrooms, the desks are set up so that students are in groups but if students are working independently while sitting at their desks (no matter how the desks are organized), you code this as "individual." If they are working in pairs, while sitting at a table with other students, code this as "partners/pairs." Similarly, if the teacher lectures to the whole class while students are sitting together at tables you would code this as whole group, regardless of the fact that they are sitting at desks that are grouped together.*

I. CLASSROOM SCAN (SCAN)		Observation Segments					
		1	2	3	4	5	6
S1	**Start Time**						
S2	**Total # of Children in Classroom**						
S3	**Total # of Adults in Classroom**						
S4	**Total # of Adults in Classroom <u>who are Interacting with Children</u>**						
S5 Classroom Structure (Code all that apply.)							
1	Whole Group/Whole Class	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
2	Large Group (more than half the class)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
3	Small Group (half or less than half the class)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
4	Centers	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
5	Partners/Pairs	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
6	Individual	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
7	Other	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐

SCAN: ITEM S6

KEY WORDS AND PHRASES		
S6	***Activity***	*We use the term "activity" in SCAN as a broad category that includes any type of classroom event, from the general (such as students having meals, or playing during center time) to the instructional (such as teacher reading to students, teachers and students discussing the science lesson, students rotating around learning centers), and everything in between. During the scan, be sure to code any type of activity that you observe.*
	Storytelling, without reading printed words	*The teacher or student uses the illustrations to tell a story without reading the actual words. This may take the form of a picture walk, or the book may be a picture book and the teacher simply makes up a story based on the pictures.*
	Whisper Read	*Students read quietly to themselves, whispering the words as they read.*
	Working with alphabet, sounding out letters/words, rhyming	*This includes times when the teacher is building students' awareness of the rhythm and sound of language, such as asking students to sound out letters or words, to recognize/ name certain letters, and/or using rhythm to help students be aware of sounds, letters and words. Or when children are playing with objects that expose them to letters, numbers, sounds, etc.*
	Vocabulary	*Students could be practicing vocabulary by going through lists of words and their meaning, practicing words on flash cards (word on one side; definition on the other side), or otherwise discussing word meaning.*
	Preparing to write	*This includes times when students are drafting stories, sharing what they have written with the teacher or peers, outlining what they will write, or otherwise preparing to write.*
	Mathematics	*This includes shapes and patterns, numbers, counting, weighing/measuring, using math operations (addition, subtraction, multiplication, division).*
	Fine Motor Play	*Playing with small objects such as beads (stringing them on a string), Legos, small blocks, puzzles, or using scissors counts as fine motor play – activities that involve moving the hands/fingers. Do not code fine motor play if students are writing, doing art, typing, or operating a mouse.*
	Gross Motor Play	*Playing with larger objects (lifting, stacking, carrying) or dancing, jumping, running, playing ball – activities that involve moving the body as a whole.*
	Dramatic and Creative Play	*Make-believe, "playing house," students may dress up, pretend to have grown-up jobs, may play with puppets or dolls, toy animals, toy cars or figurines.*
	Morning Meeting and Calendar Time	*Morning meeting/calendar time usually takes place near the beginning of the day (hence "morning"); may include one or more of the following: Question of the Day, a sentence students are asked to correct, sharing (what you did over the weekend, what you did on vacation, your favorite kind of food), survey questions (How many students are wearing red?), singing songs (particularly welcome or good morning songs), show and tell, or a Morning Message. Calendar time often involves questions about the day of the week, the month, what the date is, seasons, weather, and counting activities (such as counting days by the 5s, 10s, or 20s).*

FAQs		
S6	***Do I need to mark down what every child is doing during the scan or do I just need to mark the activities of the students working with the teacher?***	*You are supposed to account for all children when conducting the activity scan. Watch the room for 1 minute and record all of the different activities in the room.*
	What if a student moves from one activity to another while I'm conducting my scan?	*You should record all activities you observe during that 1 minute, so if a student moves from one activity to the next during your scan, you would record both of the activities you observed the student do.*
	What if I am able to scan the room in less than one minute?	*Even if you are able to account for all of the activities in the room in less than one minute, continue watching for the full minute in case any students move to a new activity.*
	Is my scan supposed to count as part of my 15 minute observation segment or in addition to it?	*You are supposed to conduct your counts and scan in one minute and then start your 15 minute observation segment; therefore you will typically have an end time that is about 16 minutes after the start time you recorded in S1.*
	What if I conduct my scan and observe for a few more minutes but the students leave to go to recess, should I code this observation segment?	*No, your 1 minute scan does not count towards your 15 minute segment. Thus, if you were only able to observe for a few minutes, erase the start time and SCAN information collected. When the children return from recess, begin another scan and observation segments. You will know from the class schedule when students are scheduled for lunch, recess and special classes (that involve them leaving the classroom, such as music or art).*

I. SCAN (continued)		Observation Segments					
		1	**2**	**3**	**4**	**5**	**6**
S6	**Type of Activity (Code all that apply)**						
Language Arts/Reading							
1	Looking at or talking about books/texts/pictures/charts/posters	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
2	Books read out loud	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
3	Storytelling, without reading printed words	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
4	Reading independently (silently or whisper reading)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
5	Working with alphabet, sounding out letters/words, rhyming	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
6	Vocabulary (word meaning)	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
7	Beginning to write letters or words, copying, tracing	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐
8	Writing sentences and paragraphs (or preparing to write)	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐
Mathematics							
9	Math activity/numbers/counting/operations/shapes/patterns/measuring	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐
Fine and Gross Motor Play							
10	Playing/working with small objects–puzzles, beads, using scissors	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐
11	Playing/working with larger objects or activities using the body–playing with large blocks, dancing, jumping, playing ball	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐
12	Playing in areas with word labels (i.e., word labels on objects)	12 ☐	17 ☐	17 ☐	17 ☐	17 ☐	17 ☐
Science/Nature and Social Studies							
13	Science activity, water table, sand or rice table; experiments or science concepts	13 ☐	19 ☐	19 ☐	19 ☐	19 ☐	19 ☐
14	Social studies activity, civics, geography, history; famous people, jobs, members of a community, places in town/school	14 ☐	14 ☐	14 ☐	14 ☐	14 ☐	14 ☐
Arts/Music/Dramatic and Creative Play							
15	Arts and music – drawing, painting, singing/music	15 ☐	15 ☐	15 ☐	15 ☐	15 ☐	15 ☐
16	Dramatic and creative play–dressing up, playing house or jobs, toy cars, puppets, animals/people	16 ☐	16 ☐	16 ☐	16 ☐	16 ☐	16 ☐
Routines/ Down Time							
17	Nap/snack/meal/bathroom/water breaks/transitions	17 ☐	17 ☐	17 ☐	17 ☐	17 ☐	17 ☐
18	Wandering (unoccupied)	18 ☐	18 ☐	18 ☐	18 ☐	18 ☐	18 ☐
Other							
19	Morning meeting/calendar time	19 ☐	19 ☐	19 ☐	19 ☐	19 ☐	19 ☐
20	Assessment/testing	20 ☐	20 ☐	20 ☐	20 ☐	20 ☐	20 ☐
21	Using computers/smart technology	21 ☐	21 ☐	21 ☐	21 ☐	21 ☐	21 ☐
22	Television/videos	22 ☐	22 ☐	22 ☐	22 ☐	22 ☐	22 ☐
23	Sharing or show and tell	23 ☐	23 ☐	23 ☐	23 ☐	23 ☐	23 ☐
24	Other	24 ☐	24 ☐	24 ☐	24 ☐	24 ☐	24 ☐

NOW SWITCH TO NOTE-TAKING BOOKLET & OBSERVE FOR 15 MORE MINUTES TAKING NOTES

II. DESCRIPTION OF SEGMENT (DESC): ITEMS D1-D4

KEY WORDS AND PHRASES		
D1-D4	***Observation segment***	*This refers to the 15-minutes when you took notes about how the teacher is interacting with the students. Ideally, the segment will be 15-minutes long. Due to last-minute changes in the classroom schedule and/or unexpected interruptions, an observation segment may be cut short, if the students need to leave the classroom. If this happens, as long as you have observed for at least 5 minutes, you can code the items in dimensions DESCRIBE through HIGH.*
D2	***# of Children Observed During Segment***	*# of children observed during segment is equal to the number of children that you included in your codes during your observation segment. This may be different than the total # of children if the class splits off into groups and you need to follow one teacher. You would then count the number of students you observed during the segment in D2.*
D3	***# of Adults Observed During Segment***	*# of adults observed during segment includes the number of adults on whom you focused during the segment. This usually will be the lead teacher. If, during the segment, two adults co-taught, then you would record two for D3.*
D4	***Special events or unusual circumstances***	*If anything unexpected occurs during the observation segment (such as a fire drill, a child or teacher getting ill so that it disrupts the class, an unscheduled visit from the principal that interrupts instruction, etc), record it in D4. Also, if you observed a person other than the lead teacher make a note of that in D4.* *If you do not experience any unusual circumstances during your observation segment you can leave it blank or write "none".*

FAQs		
D1	***Is my scan supposed to count as part of my 15 minute observation segment or in addition to it?***	*You are supposed to conduct your counts and scan in one minute and then start your 15 minute observation segment; therefore you will typically have an end time that is 16 minutes after the start time you recorded in S1.*
D2	***What is the difference between the "Total # of Children" (S2) & "# of Children Observed" (D3)?***	*Total # of children is equal to all children in the classroom. You collect this number during the 1 minute scan.* *# of children observed during segment is equal to the number of children that you included in your codes during your observation segment. This may be different than the total # of children if the class splits off into groups and you need to follow one teacher. You would then count the number of students you observed during that segment in S7.*
D3	***What is the difference between the "Total # of Adults" (S3) & "# of Adults Observed" (D4)?***	*Total # of adults is equal to all the adults in the classroom during your on 1 minute scan (including parents, other school personnel.* *# of adults observed during segment refers to the number of adults on whom you focused during the segment. This usually will be the lead teacher. If two adults co-taught during the segment (teaching together) then you would record a two for D3.*
D4	***What types of things should count as "special events" or "unusual circumstances" for item D4?***	*Use your best judgment. This is a place to record any events or circumstances that seem out of the ordinary (such as a fire drill, a child or teacher getting ill so that it disrupts the class, an unscheduled visit from the principal that interrupts instruction, etc). Also, if you observed a person other than the lead teacher make a note of that in D4.* *If you do not experience any unusual circumstances during your observation segment you can leave it blank or write "none".*

II. DESCRIPTION OF SEGMENT		Observation Segments					
		1	2	3	4	5	6
D1	End Time of Observation Segment						
D2	# of Children Observed During Segment						
D3	# of Adults Observed During Segment						

D4	Record any special events or unusual circumstances that indicate that the day was not typical.	
	SEGMENT 1	
	SEGMENT 2	
	SEGMENT 3	
	SEGMENT 4	
	SEGMENT 5	
	SEGMENT 6	

TEACHER'S USE OF LANGUAGE (LANG): ITEMS L1-L7

KEY WORDS AND PHRASES		
L1-L2	***Use of language***	*By language we mean both spoken and written language.*
L1	***Adding more information to what the student said***	*A teacher can do this by expanding on what a student says, often to clarify or strengthen relationships or encourage the student to extend his or her thinking. For example, a student says "I think the story takes place on an island because of all the water." and the teacher says, "Right, the reader knows that the setting is an island because the man described how he followed the beach all the way around and found it was surrounded by water." Note that in younger grades, teachers often repeat what a student says as part of their expansion; for example a student says "I ran to the playground!" the teacher might say "You ran to the big playground". This counts as expansion because she added to what the child said.*
	Narrating teacher or student actions (self or parallel talk)	*Teachers may narrate their own actions or a student's actions. For example, a teacher might narrate what they are doing as they make a painting: "Look at how I am painting my picture; I'm mixing the colors up, putting the paint on my paintbrush and painting swirls on the paper." or if a student is building something with blocks the teacher might say "I like how you are building that tower; I see that you are stacking the little blocks on top of the big blocks." Do not include forecasting ("Now, I'm going to read this book") as narrating teacher or student actions.*
	Open-Ended Questions	*Questions that require more than a one-word answer or a Yes/No response. Open ended-questions often require the student to expand their language.* "*Tell me more," "And then what?" "What else did he do?"* are typical open-ended questions. *"Do you feel happy?" is not an open-ended question; "Why do you feel happy?" is.* *Yes/No questions are not open-ended, nor are times when students use a complete sentence to give a one-word answer: What is this? It's a penny.*
	Time to Respond	*The teacher does not interrupt students, but allows time for students to respond to questions (about 3-5 seconds) so that students have time to think before they speak*
	Complete Sentences	*When students give a brief response to a question (when speaking or writing), the teacher may ask them to use a complete sentence, to encourage them to use more language.*
L4	***Purpose of teacher's talk***	*For this item you have to determine what the teacher(s) you are observing most often communicated when they spoke (behavior management, giving instructions or directions, teaching, or social conversing with the students). Sometimes, it may be hard to pick the most frequent purpose. If it is a toss-up between two or more, choose the lesser one (i.e. if it was about 50% behavior management and 50% giving directions you would code "2" behavior management).*
L5	***In-depth Conversations***	*This item captures whether an in-depth conversation occurs by counting the number of turns a single student has with a teacher on a topic. The conversation can be started by the teacher or the student; as long as the student gets three uninterrupted turns, it is considered an in-depth conversation.* *For example, the teacher asks a student a question, the student answers, the teacher asks the student to explain his or her answer, and the student explains his or her answer, the teacher probes again and the student answers a third time (T -> S ->T ->S ->T->S). If however, the teacher turns and speaks to another student or adult at any time in between the three student turns it would not count as an in-depth conversation.* *or the reverse where the student says something to the teacher, the teacher responds, and the student says a second thing, the teacher responds and the student says a third thing (S -> T ->S ->T ->S).* *L1 captures ways the teacher encourages students to speak more, to use more language. L5 captures whether the teacher probed a single topic in-depth with a student.*

FAQs		
L1-L8	***What if there's more than one adult in the classroom who talks?***	*If two adults are leading the class together – i.e., co-teaching – treat them as if they are one adult and code their language as if they were one person. If they split into groups, follow the general rules for whom to observe and code only for that teacher's language. See observation procedures (pages 1-2) for more information on the rules for who to observe.*
L2	***Do we code the number of times a technique is used even if it's the same technique?***	*Yes, if the teacher uses the same technique 3 times during the observation segment, you code it as a "3." If the teacher uses three different techniques during the observation segment, you code it as a "3."*
L5	***Should I code conversations between the teacher and students if they take place only during a lesson?***	*No. You should code any and all conversations/exchanges between the teacher and students you are observing. They can take place at any point in the observation segment.*
	Should I code conversations that take place between students?	*No. Items L1-L8 are only asking about the teacher's language or the students' language when talking to the teacher. You are not coding for student conversation here.*
	Does a one word answer count as a "student turn"?	*Yes, you are not responsible for coding the quality of the conversation; you are coding whether one or more students had a conversation with a teacher.* *Similarly, if a student responds with something like "I don't know" that counts as a turn.*
L6	***What if the teacher speaks in an language other than English, does this count as not clear?***	*No, focus on the clarity of the teacher's use of English.*
L6	***Do mispronunciations and misspellings count as grammar mistakes?***	*No. Only count grammatical errors , such as double negatives, agreement errors (The boy go to the store), verb tense errors (Tomorrow, I played soccer.), or "ain't."*

III. TEACHER'S USE OF LANGUAGE (LANG)			Observation Segments					
			1	2	3	4	5	6
L1	**What techniques did the teacher use to help students expand their use of language?** ***(Code all that apply.)***	1 No talk/no encouragement to expand language	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
		2 Teacher added more information to what the student said.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
		3 Teacher narrated student actions or his or her own actions.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
		4 Teacher asked open-ended questions or questions that help students say more.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
		5 Teacher allowed students time to respond to questions (about 3-5 seconds).	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
		6 Teacher and/or students sang songs or recited poems.	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
		7 Teacher asked students to use complete sentences or to use a word in a sentence.	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐
L2	**How often did the teacher use a technique from L1 to help students expand their use of language?** ***(Code only one.)***	1 No talk/no techniques used.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
		2 Technique(s) used 1-2 times.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
		3 Technique(s) used 3-4 times.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
		4 Technique(s) used 5 or more times.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
L3	**How much of the time did the teacher do things other than talk with the students?** ***(Code only one.)***	1 Never or almost never (Teacher spoke with the students for almost the whole segment.)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
		2 Sometimes (Teacher spent less than 5 minutes doing things other than talking to students.)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
		3 Most of the time (Teacher spent 5-10 minutes doing things other than talking with students.)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
		4 Almost all or all of the time (Teacher spent more than 10 minutes doing things other than talking with students.)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
L4	**What was the most frequent purpose of the teacher's talk?** ***(Code only one.)***	1 The teacher did not speak.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
		2 The teacher's talk was mostly on behavior management.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
		3 The teacher's talk was mostly giving directions.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
		4 The teacher's talk was mostly for social conversation.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
		5 The teacher's talk was mostly for instruction or content.	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
L5	**Did the teacher have an in-depth conversation with a single student on a topic?** ***(Code only one.)***	1 No in-depth conversation between teacher and any student	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
		2 A single in-depth conversation	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
		3 More than one in-depth conversation	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
L6	**How clear and distinct was the teacher's speech?** ***(Code only one.)***	1 The teacher did not speak.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
		2 At times the teacher's speech was unclear: words were slurred together, pronunciations distorted and/or pacing was too fast.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
		3 The teacher's speech was mostly clear and easy to understand.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
L7	**Did the teacher make any grammar mistakes when speaking?** ***(Code only one.)***	1 The teacher did not speak.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
		2 Yes, the teacher made a grammar mistake.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
		3 No, the teacher did not make any grammar mistakes.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐

TEACHER'S USE OF LANGUAGE (LANG): ITEM L8

KEY WORDS AND PHRASES		
L8	***Descriptive, sophisticated or technical words***	*Descriptive words are words that go beyond providing a simple name by offering more information about an object or subject. Listen for specific nouns (rather than common nouns or pronouns: "Please pick up the football." vs. "Please pick the ball up.[or Please pick that up.]"), verbs (i.e. ambled, jogged), adjectives (cherry red) and adverbs (briskly walking).* *Sophisticated words are words that you would not usually hear children using; they are more commonly used by adults. Instead of saying, "Please read by yourselves." the teacher might say: "Please read independently." Instead of saying, "The reason is clear." the teacher might say, "The justification is clear." Technical words are those used for specific professions or academic contexts, such as zoologist, hereditary, democracy.*

FAQs		
L8	***What if the teacher reads out loud or shows a video? Do I count that language?***	*Yes. If they use rich, descriptive words and/or sophisticated/technical words. The focus is on whether the students are exposed to rich, descriptive words and/or sophisticated/technical words, not on the origin of the words.*

L8. Did the teacher use words that are very descriptive, sophisticated and/or technical? List up to 10 words for EACH segment observed.

SEGMENT 1		
1 ☐ _____	5 ☐ _____	8 ☐ _____
2 ☐ _____	6 ☐ _____	9 ☐ _____
3 ☐ _____	7 ☐ _____	10 ☐ _____
4 ☐ _____		

SEGMENT 2		
1 ☐ _____	5 ☐ _____	8 ☐ _____
2 ☐ _____	6 ☐ _____	9 ☐ _____
3 ☐ _____	7 ☐ _____	10 ☐ _____
4 ☐ _____		

SEGMENT 3		
1 ☐ _____	5 ☐ _____	8 ☐ _____
2 ☐ _____	6 ☐ _____	9 ☐ _____
3 ☐ _____	7 ☐ _____	10 ☐ _____
4 ☐ _____		

SEGMENT 4		
1 ☐ _____	5 ☐ _____	8 ☐ _____
2 ☐ _____	6 ☐ _____	9 ☐ _____
3 ☐ _____	7 ☐ _____	10 ☐ _____
4 ☐ _____		

SEGMENT 5		
1 ☐ _____	5 ☐ _____	8 ☐ _____
2 ☐ _____	6 ☐ _____	9 ☐ _____
3 ☐ _____	7 ☐ _____	10 ☐ _____
4 ☐ _____		

SEGMENT 6		
1 ☐ _____	5 ☐ _____	8 ☐ _____
2 ☐ _____	6 ☐ _____	9 ☐ _____
3 ☐ _____	7 ☐ _____	10 ☐ _____
4 ☐ _____		

TIME (TIME): ITEMS T1-T3

KEY WORDS AND PHRASES		
T1-T3	***Teaching Time***	*All classroom time is potential teaching time. Teaching time counts as any time the lead teacher could be interacting with at least one student to help them learn. Interaction includes talking with students and/or monitoring students as they work/play independently or in groups. If the students are present and the teacher is not interacting with them because he or she is dealing with student disruptions, transitions or down time, or doing tasks other than interacting with students, that is considered "lost teaching time".*
T1	***Time lost due to student disruption***	*This is time when a teacher stops teaching to deal with student misbehavior. This might include a teacher reprimanding a student for talking to friends, throwing paper, refusing to follow directions, refusing to sit down or sit still, touching friends, and/or misusing supplies. Also, include student disruption that causes the teacher to delay the daily routine, for example, student misbehavior that prevents students from lining up or sitting down in a circle.* *For this item, only count the amount of time that the teacher stops teaching or halts the routine because of student disruption.*
T2	***Time lost due to transitions***	*Count time spent changing from one activity or lesson to another (from writing time to math time, from reading to writing), or moving from one place in the room to another (e.g., from the carpet to tables or desks, from the calendar area to the carpet) without any instruction. Transitions are NOT always physical movement but do signal a change in topic, focus, or activity. You can consider a transition time over once the teacher is working with at least one student again.* *Note: Some teachers use transition time for instruction, for example having pre-kindergarteners count as they line up to go to the bathroom. When calculating the time lost due to transitions, do not include any time that the teacher provided instruction or learning during a transition.*
	Time lost due to down time	*This includes time spent waiting for the teacher or students to prepare for an activity (getting out notebooks, crayons, folders, books, getting meals, tending to a sick child or any other non-instructional issue) and any time spent going to the bathroom or water fountain as a whole class.*
T3	***Tasks other than interacting with students***	*Teachers may grade papers, prepare for the next activity, or talk to other adults. Count any time the teacher is engaging in activities that do not involve interacting with/monitoring students.*

FAQs		
T1	***How do I keep track of how much teaching time was lost due to disruptions from students?***	*To help with coding item TM1, make note every time the teacher has to stop teaching to address student behavior. Count the seconds/minutes until teaching begins again. At the end of the observation segment, add these times together to code item TM1.*
T2	***How do I keep track of how much teaching time was lost due to transitions/ down time?***	*To help with coding item TM2, make note every time the teacher and students begin to move from one place, lesson, or activity to another. Count the seconds/minutes until the teacher begins the next lesson or activity begins. At the end of the observation segment, add these times together to code item TM2. Remember, not to count any time that the teacher provided instruction or learning for at least one student as part of a transition.*
T3	***What if the teacher is grading papers, preparing materials, talking to other adults or cleaning up (etc.) and is NOT engaged with at least one student while the students are present in the classroom?***	*This counts as lost teaching time and should be coded as part of TM3. Even if the students are working independently or with their peers while the teacher engages in these other activities, it should be considered lost teaching time.*

IV. TIME (TIME)		Observation Segments					
		1	2	3	4	5	6
T1 How much teaching time was lost due to disruptions from students? ***(Code only one.)***	1 Rarely or not at all (less than 1 minute)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 A little time (1 - 2 minutes)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Some of the time (more than 2 but less than 5 minutes)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Most of the time (5 to 10 minutes)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 All or almost all of the time (more than 10 minutes)	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
T2 How much teaching time was lost due to transitions/ down time? ***(Code only one.)***	1 Rarely or not at all (less than 1 minute)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 A little time (1 - 2 minutes)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Some of the time (more than 2 but less than 5 minutes)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Most of the time (5 to 10 minutes)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 All or almost all of the time (more than 10 minutes)	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
T3 How much teaching time was lost because the teacher was doing tasks other than interacting with the students? ***(Code only one.)***	1 Rarely or not at all (less than 1 minute)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 A little time (1 - 2 minutes)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Some of the time (more than 2 but less than 5 minutes)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Most of the time (5 to 10 minutes)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 All or almost all of the time (more than 10 minutes)	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐

ENGAGEMENT (ENG): ITEMS ENG1-ENG5

KEY WORDS AND PHRASES		
E1	***Engagement***	*For this study, engagement refers to the ways in which teachers structure instructional activities to engage students. This can include listening or reading, answering or asking questions, playing games, drawing, acting, or other forms of engagement.*
	Teacher asked students questions	*This includes times when the teacher asks the whole class questions as well as asking individual students questions.*
	Choral reading/response	*This includes times when the teacher asks students to read along as he or she reads a text out loud. This could include reading only one word at the end of each page, or reading all of the words on the page. Teachers may invite students by asking them specifically ("Read along with me.") or my gesturing (open hands, nodding) to indicate that they should read out loud too. It also includes times when the students call out words together when not reading.*
	Teacher had students speak with each other	*This includes any time when the teacher directs students to talk with each other.*
	Teacher had students engage in a hands-on activity	*This includes activities such as conducting experiments, building houses, or painting a mural. (Writing and drawing does not count as a hands-on activity.)*
	Writing	*This includes times when the students are writing and also times when the students are composing and the teacher writes for them.*
E2	***Enthusiasm***	*Teachers demonstrate enthusiasm by the tone of their voices (excited); by their gestures, facial expressions and posture; and by what they say about what they are teaching (why what students are doing is important, how it is important) and by showing an interest in students.*
E3	***Call on students***	*This includes any way that the teacher uses to indicate whose turn it is to speak. The most common is when teachers name the student who should speak, point at the student, or face the student and nod.*
E5	***Briefly discuss with peers***	*This includes any time students are asked to speak with each other for a short time (4 minutes or less) (partners, small groups). Common approaches include pair/share, whisper to friend, and turn to your partner.* ***Whisper to a friend*** *is just that – the teacher asks students to quickly whisper the answer to a question to their neighbor.* ***Turn to your partner*** *and* ***pair/share*** *is the same as whisper to a friend, without the whispering.*
	Discuss with peer(s) for more than 4 minutes.	*This includes any time students are asked to speak with each other for more than 4 minutes. Think/pair/share is a common name for this type of activity.* ***Think/pair/share*** *is a type of partner work where the teacher gives students a question or topic to think about. Then, the teacher pairs students (or has students get into pairs) and share their thoughts/ideas.*
	Work with peer(s) on a hands-on activity.	*This includes activities where the teacher has students work together to conduct experiments, build houses, paint a mural. (This does not include drawing, which is coded for option #5 for E1.)*
	Informal student interactions/talk	*This includes conversation among students that the teacher allows, but is not directed by the teacher.*

FAQs		
E1	***What if the teacher is working with 6 students in a group, and 3 students answer all of the questions, while the other 3 just listen?***	*You are supposed to code all the different types of engagement in E1.Given this example you would code "Students listened and/or read" and "Students answered or asked questions orally".* *Please note that option 1: Did not engage one or more students in activities, does not mean you need to watch for or code off task behavior. This item looks at the ways teacher engages students. Option one is used for wandering or unoccupied students.*
E1, E2, E3, E4, E5	***How do I code if I observe an activity that is NOT whole group (i.e., small group or individual)?***	*Using the general coding rules you would only be coding the engagement items for the students you observed interacting with the teacher you are following. For example, if the teacher is working with a small group of four students while the rest of the class does work independently, you would observe the small group and code engagement for those four students. If the teacher also disciplines two students who are supposed to be working independently but are talking to one another you would include those two students in your codes.*
E1 E3 E4	***What if the teacher and students talk together, such as if they recite a poem or read something together out loud? OR What if the students recite a poem for the teacher, and the teacher doesn't talk?***	*This type of choral reading or poem recitation would be coded as a "5" for E1 and as a "5" for E3, because the teacher called on all the students to speak when he or she invited them to choral read.* *But for E4, we do not count choral responses or poem recitation as times when the individual students spoke with the teacher (because everyone is speaking together, individual students aren't speaking directly to the teacher).*
E2	***How do I code E2 if the teacher showed enthusiasm once or twice during a segment but was generally more interested and focused than enthusiastic?***	*For E2 you are supposed to code the overall level of enthusiasm so given the fact that the teacher only demonstrated enthusiasm once or twice during that segment, you should code "Somewhat; appeared focused but not enthusiastic"*
	What if a teacher seems to show enthusiasm but it's hard to know if it's genuine?	*If the teacher demonstrates enthusiasm you do not need to judge how genuine it is; you can code it as enthusiasm.*
E3	***Does the teacher need to explicitly call on students to count in E3?***	*Yes, the teacher must specifically call on a student by saying his or her name, or otherwise gesturing to the student so that he or she know it is his or her turn. If the teacher directs the question to the whole class for them to all answer chorally then count it as "Almost all or all". If a teacher asks a question and a student calls out the answer, this does not count as the teacher calling on that student, regardless of whether the teacher accepted the student's answer.*
E5	***What about in center time when they may work with another student for a part of an activity and then go off and do something on their own?***	*For E5, check all types of interactions students had during the observation segment.*

V. ENGAGEMENT (ENG)			Observation Segments					
			1	2	3	4	5	6
E1 In what ways did the teacher engage students in activities? *(Code all that apply.)*	1	Did not engage one or more students in activities.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2	Students listened to the teacher and/or read silently.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3	Teacher asked students questions.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4	Teacher had students play games.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5	Teacher had students draw, act, sing and/or invited them to read along (choral reading/response).	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6	Teacher had students speak with each other.	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
	7	Teacher had students engage in a hands-on activity.	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐
	8	Teacher had students write numbers, letters, words, phrases (less than sentences).	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐
	9	Teacher had students write a sentence or more.	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐
	10	Teacher had students write about the topic, characters and/or ideas in a book/text.	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐
	11	Teacher had students use a book/text as a model for their writing.	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐
	12	Students were actively engaged in some other way.	12 ☐	12 ☐	12 ☐	12 ☐	12 ☐	12 ☐
E2 Overall, how enthusiastic was the teacher? *(Code only one.)*	1	Not at all; appeared bored or disinterested	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2	Somewhat; appeared focused but not enthusiastic	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3	Very; appeared enthusiastic and highly interested	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
E3 How many different students did the teacher call on during the segment? *(Code only one.)*	1	Almost none or none of the students (2 or fewer)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2	Less than 1/2 the class or group observed	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3	About 1/2 the class or group observed	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4	More than 1/2 the class or group observed	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5	Almost all or all of the students observed	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
E4 How many individual students spoke with the teacher? *(Code only one.)*	1	None of the students spoke with the teacher	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2	Less than 1/2 of the class or group observed	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3	About 1/2 of the class or group observed	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4	More than 1/2 of the class or group observed	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5	Almost all of the students or group observed	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
E5 In what ways did the teacher encourage student interaction? *(Code all that apply.)*	1	No interactions with other students.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2	Teacher had students read with partners.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3	Teacher had students briefly discuss with peer(s) (4 minutes or less).	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4	Teacher had students discuss with peer(s) for more than 4 minutes.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5	Teacher had students work with peer(s) on a hands-on activity.	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6	Teacher had student(s) lead an activity.	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
	7	Teacher allowed informal student interactions/talk.	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐
	8	Other	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐

BOOK OR TEXT SHARING (READ): ITEMS R1-R10

KEY WORDS AND PHRASES		
R1-R10	***Book/Text***	*All bound books: story books, textbooks, biographies, history books, picture books (that the teacher reads as a story) or other reading material: photocopies or handouts; magazines, newspapers, brochures, student writing, and teacher writing (of at least one sentence in length)* *NOT considered book/text: Lists posted on the walls or boards, such as classroom rules, schedules, directions or assignments.*
R1	***Observed the beginning of a book/text sharing***	*Code this if you observed a pre-reading activity that clearly relates to a specific book/text OR if you observed the beginning of a book/text reading.*
	Observed the book/text reading	*Code this if you observed the actual book/text reading. This includes students reading independently or in groups, students being read to, a teacher verbalizing a picture book, and discussion about the text (while reading/story telling is in progress).* ***Emergent Reading:*** *If you observe a teacher reading a book/text with repeatable patterns, rhymes, or songs, and the students join in by labeling pictures, calling out words or parts of the text, or "singing along," this counts as book/text reading.*
	Observed the end of a book/text sharing	*Code this if you observed the end of a book/text sharing. The teacher may not have finished reading a book in one sitting. You do not need to see the end of the book, just the end of the reading activity.*
R2	***Read out loud***	*This includes times when the teacher reads a text out loud to the students or when he or she plays an audio tape or CD with someone else reading the text out loud.*
R4	***Whisper-Read***	*When children read quietly to themselves, whispering the words as they read.*
R7	***Predictable texts***	*Includes texts where lines are repeated regularly, such as: "(on one page) Brown Bear, Brown Bear, what do you see? I see an eagle looking at me; (on the next page) Brown Bear, Brown Bear, what do you see? I see a camel looking at me."*
R10	***Presentation Techniques***	*When reading out loud, teachers may change their style of speech (the use of tone of voice, changes of volume, pacing, using different voices and/or use of body language,(gestures, and facial expressions) or props such as using pictures, puppets, or models to demonstrate the setting of the story or what is taking place.*

FAQs		
R1	***What if there isn't any book/text sharing during the Observation Segment?***	*Check "1" (No reading activities) for item R1, draw a line down the READ items in that segment, and proceed to the next section, VOCAB.*
	What if I only see one or two aspects of reading (e.g., just reading and post-reading) during the observation segment?	*That's okay. Code only what you see during the observation, not what you think might have happened before or what might happen later.*
	What if the teacher introduces a book in the morning that they will read in the afternoon or on another day?	*Code only what you see during the observation segment. If the teacher introduces a book but does not read or have students read the book during the segment, then code for pre-reading but not for during reading.*
R1-R10	***How do I code if I observe an activity that is NOT whole group (i.e., small group or individual)?***	*If the teacher works with small groups or individual students, follow the teacher as he or she moves from group to group or from student to student. Code what you observe with all groups/students the teacher worked with.*
R2	***What if someone other than the teacher reads to the students?***	*You would focus on what the teacher is doing during this time. If the teacher you are focusing on is not involved in the book reading, then you code what the teacher is doing. If the teacher joins the class and the visiting reader, then you would code the activity in which the teacher is involved.*
R5	***Students didn't read books/texts***	*If the teacher reads out loud to the students (or even to just one student) and the students are not reading along, out loud, with the teacher, we view this as "students didn't read books/texts." During read aloud, we do not infer that students are reading the words on the page unless we hear them doing so.*
R7	***How can you tell what kind of book it is, just from observing?***	*Pay attention to how the teacher describes the book (i.e., "Today we're going to learn about how bread is made."). Pay attention to the content of the book, if it's read out loud.*
R7, R8	***How can you tell what type of book it is if students are reading different books?***	*Pay attention to anything the teacher might say about the books (i.e., "Oh Josh, I see you chose another fiction book.") If you can see the titles of books and can tell by the title, code accordingly.*
R8	***What if the teacher reads more than one book, and one is a picture book but the other has one sentence on each page? How do I code?***	*You would check a "2" for only pictures and a "4" for one or two sentences on most pages.*
	How can I tell how many words are on the pages of the texts the teacher and/or students are using?	*If the teacher reads the book/text out loud and/or holds the book up so that you can see the words on the pages, it is easy to tell. If not, do your best to estimate how many words are on the pages of the book the teacher is reading to the class. If students are reading, try to see the books of the students sitting near you and estimate the number of words as best you can.*
	How do I code R8 (How many words were on the pages?) when more than one book/text is used during the segment?	*This item is "Code all that apply." Record the number of words on the pages of any and all books observed during the segment.*
R9	***How do I code the number of books the teacher introduced/read/ discussed if all the students are reading different books?***	*You only need to code the number of books/texts that the teacher introduces, reads or discusses. If the students are all reading independent books and the teacher does not introduce/read/or discuss any of the books with any students, then code 0.*

VI. BOOK OR TEXT SHARING (READ)		Observation Segments					
VI A. CLASSROOM CONTEXT FOR BOOK/TEXT SHARING		**1**	**2**	**3**	**4**	**5**	**6**
R1 Which parts of book/text sharing did this segment include)? *(Code all that apply.)*	1 No reading activities (DRAW LINE DOWN SEGMENT & GO TO VOCAB)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Observed pre-reading activity or beginning of book/text reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Observed the book/text reading	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Observed the end of a book/text sharing	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
R2 Did the teacher share a book/text with the students? *(Code all that apply.)*	1 No, did not read book/text or tell a story	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Yes, read out loud	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Yes, made up a story based on the pictures in the book	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
R3 Who did the teacher read to? *(Code all that apply.)*	1 Did not read the book/text	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 The entire class	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Small groups or individuals	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
R4 How did the students, who worked with the teacher, read books/texts? *(Code all that apply.)*	1 Students didn't read the books/texts	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Out loud (individual students reading aloud)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Silently (individual students reading silently)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Whisper-read (each reading quietly to themselves)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Choral reading (students reading together out loud)	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Students completed an assignment while reading	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
R5 Did the student(s), who worked with the teacher, read the same books/texts as the other students? *(Code all that apply)*	1 Students didn't read books/texts.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Only one student worked with the teacher, so only one book.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Students read the same book.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Students read different books.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
R6 Who did students, who worked with the teacher, read with or to? *(Code all that apply.)*	1 Students didn't read books/texts	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 By themselves, independently	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Individually, with the teacher	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 In pairs or small groups, without the teacher	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 In pairs or small groups, with the teacher	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 With or to the whole class	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
R7 What types of books/texts did the teacher and/or students read (or prepare to read/preview)? *(Code all that apply.)*	1 Did not read books/texts	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Books/texts that tell a story	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Books/texts that present information	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Books/texts that tell a story and present information	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Books/texts that include poems, songs, and predictable texts with repeated lines	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Books/texts that tell "how to" do something	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
	7 Text related to morning meeting (sentence of the day)	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐
	8 Students' writing	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐
	9 Teachers' writing	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐
	10 Other	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐
R8 How many words were on the pages of the books/texts read by the teacher (not including student or teacher writing)? *(Code all that apply.)*	1 Did not read books/texts, or texts were produced by student(s) or teacher	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No words, only pictures (or a few words on 1-2 pages)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 One word or a few words on most pages	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 One or two sentences on most pages	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 A paragraph or two on most pages	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Chapter books	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
R9 How many books/texts did the teacher introduce/read/ discuss? Record #							
R10 When reading out loud, what did the teacher emphasize with presentation techniques (style of speech, body language)? *(Code all that apply.)*	1 Teacher didn't read books/texts out loud	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Teacher did not use presentation techniques	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Things **other than the content** or subject of the text (such as letters, word sounds, rhymes, rhythm, sentence structure)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Things **related to the content** or subject of the text (such as characters' emotions, story tone, characters' voices, information from the text)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐

FOCUS ON MEANING OF BOOK/TEXT: ITEMS R11-13

KEY WORDS AND PHRASES		
R11 R12 R13	***Pre-reading***	*Pre-reading is how the teacher introduces the book/text, including any type of activity, discussion, or teacher talk that prepares students to read the book/text. Must precede the reading itself and must relate to the act of reading a specific book/text.*
R11	***Announces beginning of reading activity***	*This is when the teacher provides a general announcement that they are starting a reading activity. For example s/he might say "It's time for reading" or "Everyone get out their book of the day for silent reading time".*
	Letters or words (sounding out letters or words; rhyming words; word recognition)	*Teachers may focus students' attention on letters and/or words.* *You may see two main approaches teachers use to help students recognize words: decoding and sight word recognition. First, decoding is when we sound out words that don't look familiar to us, to see if we recognize them. Teachers use many techniques to help develop students' abilities to sound out words, beginning with sounding out letters, as part of a reading activity or independently.* *Second, sight words are "high frequency" or common words such as and, the, to, my, that, I, because, your, why, until, and first. Students are taught to recognize words by sight, rather than by sounding them out. Teachers will often practice sight words at morning meeting or during the sentence of the day.*
	Grammar/Mechanics/Spelling	*Grammar refers to how we construct sentences, types of sentences and parts of speech, and all the rules involved (such as agreement, verb tenses). Mechanics includes capitalization and punctuation. For example, a teacher may write a few sentences on the board for morning meeting, intentionally leaving out the periods at the end of the sentences, and then ask students questions about what needs to be fixed (add punctuation).*
	Key Features of the Text	*This includes the type of text it is (fiction/non-fiction; fable, adventure, science fiction, etc); and the authors and illustrators; the physical parts of the book: front cover, title page, beginning of text, end of text, back cover.*
	Text Structure	*Recognizing and using the author's organizational plan or text structure to help students understand and remember what they have read. Common elements of text structure include that each story has a beginning, middle and end, that there are main characters, characters' goals, a setting, key events, problems and solutions. Teachers often present a chart or diagram that shows the key aspects of text structure and/or have students complete a chart or diagram during or after reading. For informational text, text structure can include main ideas and supporting details; sequence of events; and problem/solution structures used to organize information.*
	Reading Comprehension strategies	*Reading comprehension strategies involve the deliberate use of a cognitive routine by the reader before, during, or after reading. These cognitive routines are specific mental actions (such as previewing, predicting, making prior knowledge connections, summarizing, self-questioning, clarifying, and visualizing) that facilitate a better understanding of text. To count as a strategy, the teacher has to identify the strategy (label it or explain it).*
	Purpose of reading the text	*Teachers may explain why he or she is having the students read a particular text. For example, a teacher might say "Today we are going to read this book to learn about dinosaurs." This does NOT include times when the stated purpose of reading is to practice a strategy – a comprehension strategy or word recognition strategy – or when a teacher says, "Let's read to find out what happens in this story." (too general)*
	Title, topic, subject and/or theme of text	*What the book or text is about, the story events or world information that one would find out by reading the text. This does not include decoding or naming single words.*
	Previewing	*The teacher explains what the book/text will be about. A common form of previewing is to do a "picture walk," where the teacher leads the students in reviewing the pictures in the book, so that they can get a sense what the book will be about. You can tell when this is different than reading a picture book because the teacher's questions are focused on what the story will be about.*
	The characters in the text, who they are, their motivation and/or goals	*This includes references to what the characters look like, who they are likely to be, how they might be feeling, what they might want to accomplish.*
	Connecting content with students' prior knowledge/experiences	*This includes times when the teacher relates content from the book/text with students' prior knowledge and/or experiences. It includes times when the teacher invites students to make these connections as well as times when the students bring up the connection ("This character reminds me of my uncle.") and the teacher affirms and/or reinforces the connection ("Good job making connections."). Connecting prior knowledge to information in books/texts includes making connections from the student's personal experience, something the student learned previously either in or out of school, another area of study, or another text or book.*
R13	***Organization when talking about the content of books/texts***	*Teachers may organize pre-reading discussions by simply announcing a general topic or subject ("We're going to read a book about George Washington. Let's make a list of what we already know about him.") Or, teachers may provide a structure to organize the discussion ("We're going to read a book about George Washington. Let's first list what we know about his childhood, then about what he did during the Revolutionary war, then what he did as president.) Teachers may organize a picture walk by asking questions that help students link what they are seeing on the different pages, such as: "Who are the characters on this page?"*

FAQs		
R11	***What if I only see one or two aspects of reading (e.g., just reading and post-reading) during the Observation Segment?***	*That's okay. Code only what you see during the observation, not what you think might have happened before or what might happen later.*
	How do I code if I observe an activity that is NOT whole group (i.e., small group or individual)?	*If the teacher works with small groups or individual students, follow the teacher as he or she moves from group to group or from student to student. Code what you observe with all groups/students the teacher worked with.*
	What if a teacher doesn't do all the talking, but asks students questions and the students do most or almost all the talking?	*This counts as talk and you should code these items.*

VI B. FOCUS ON MEANING OF BOOK/TEXT		Observation Segments					
BEFORE READING (PRE-READING)		**1**	**2**	**3**	**4**	**5**	**6**
R11 What did the teacher talk/ask about during pre-reading? *(Code all that apply.)*	1 Did not observe beginning of reading or no talk before reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Announces the beginning of the reading activity	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Letters or words (sounding out letters or words; rhyming words; word recognition)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Vocabulary (word meaning)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Grammar/mechanics/spelling	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Key features of the book/text (type of book, parts of the book, author/illustrator)	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
	7 Text structure (parts of a story/text)	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐
	8 Reading comprehension strategies	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐
	9 The purpose for reading the text	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐
	10 Title, topic, subject and/or theme of text to be read	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐
	11 What the text may be about (previewing)	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐
	12 The characters in the text, who they are, their motivation and/or goals	12 ☐	12 ☐	12 ☐	12 ☐	12 ☐	12 ☐
	13 Connecting content with students' prior knowledge/experiences	13 ☐	13 ☐	13 ☐	13 ☐	13 ☐	13 ☐
	14 Other	14 ☐	14 ☐	14 ☐	14 ☐	14 ☐	14 ☐
R12 How much detail did the teacher use when talking about the content of the book/text during pre-reading (#11, 12 or 13 from R11)? *(Code the highest.)*	1 Did not observe beginning of reading or no talk before reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No talk about content (did not code 11, 12 or 13 in R11)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Talk included 1-2 details about content	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Talk included 3 or more details about content	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
R13 How did the teacher organize the talk about content during pre-reading (#11, 12 or 13 from R11)? *(Code the highest.)*	1 Did not observe beginning of reading or no talk before reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No talk about content (did not code 11, 12 or 13 in R11)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Talked about content but only 1-2 details or details were not organized	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Talked about content and at least some of the details were organized	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐

FOCUS ON MEANING OF BOOK/TEXT: ITEMS R14-16

KEY WORDS AND PHRASES		
R14-R16	***During reading***	*During reading includes students reading independently or in groups, students being read to, a teacher verbalizing a picture book, and discussion about the text (while reading/story telling is in progress).* ***Emergent Reading:*** *If you observe a teacher reading a book/text with repeatable patterns, rhymes, or songs, and the students join in by labeling pictures, calling out words or parts of the text, or "singing along," this counts as reading.*
R14	***Talk relates to the book/text, but is not about the topic or content of the book.***	*Code this when the teacher and/or students talk about reading the book (such as when the teacher announces that it's time to read the book/text and holds up the book/text) but do not discuss the topic or content of the book. For example, "Let's go, it's time for us to read this story. Everyone in your seats."*
	Letters or words (sounding out letters or words; rhyming words; word recognition)	*Teachers may focus students' attention on letters and/or words.* *You may see two main approaches teachers use to help students recognize words: decoding and sight word recognition. First, decoding is when we sound out words that don't look familiar to us, to see if we recognize them. Teachers use many techniques to help develop students' abilities to sound out words, beginning with sounding out letters, as part of a reading activity or independently.* *Second, sight words are "high frequency" or common words such as and, the, to, my, that, I, because, your, why, until, and first. Students are taught to recognize words by sight, rather than by sounding them out. Teachers will often practice sight words at morning meeting or during the sentence of the day.*
	Grammar/Mechanics/ Spelling	*Grammar refers to how we construct sentences, types of sentences and parts of speech, and all the rules involved (such as agreement, verb tenses). Mechanics include capitalization and punctuation. For example, a teacher may write a few sentences on the board for morning meeting, intentionally leaving out the periods at the end of the sentences, and then ask students questions about what needs to be fixed (add punctuation).*
	Key Features of the Text	*This includes the type of text it is (fiction/non-fiction; fable, adventure, science fiction, etc); and the authors and illustrators; the physical parts of the book: front cover, title page, beginning of text, end of text, back cover.*
	Text Structure	*Recognizing and using the author's organizational plan or text structure to help students understand and remember what they have read. Common elements of text structure include that each story has a beginning, middle and end, that there are main characters, characters' goals, a setting, key events, problems and solutions. Teachers often present a chart or diagram that shows the key aspects of text structure and/or have students complete a chart or diagram during or after reading. For informational text, text structure can include main ideas and supporting details; sequence of events; and problem/solution structures used to organize information.*
	Reading Comprehension strategies	*Reading comprehension strategies involve the deliberate use of a cognitive routine by the reader before, during, or after reading. These cognitive routines are specific mental actions (such as previewing, predicting, making prior knowledge connections, summarizing, self-questioning, clarifying, and visualizing) that facilitate a better understanding of text.*
	Purpose of reading the text	*Teachers may explain why he or she is having the students read a particular text. For example, a teacher might say "Today we are going to read this book to learn about dinosaurs." This does NOT include times when the stated purpose of reading is to practice a strategy – a comprehension strategy or word recognition strategy – or when a teacher says, "Let's keep reading to find out what happens in this story." (too general)*
	Title, topic, subject and/or theme of text	*What the book or text is about, the story events or world information that one would find out by reading the text. This does not include decoding or naming single words.*
	The characters in the text, who they are, their motivation and/or goals	*This includes references to what the characters look like, who they are likely to be, how they might be feeling, what they might want to accomplish.*
	Connecting content with students' prior knowledge/experiences	*This includes times when the teacher relates content from the book/text with students' prior knowledge and/or experiences. It includes times when the teacher invites students to make these connections as well as times when the students bring up the connection ("This character reminds me of my uncle.") and the teacher affirms and/or reinforces the connection ("Good job making connections."). Connecting prior knowledge to information in books/texts includes making connections from the student's personal experience, something the student learned previously either in or out of school, another area of study, or another text or book.*
R13 R16	***Organization***	*Teachers may organize talk during reading by providing or eliciting a framework that organizes the information being read into categories or sequences of events ("Let's make a list of what we are learning about George Washington's childhood, then about what he did during the Revolutionary War, then what he did as president, etc.").Teachers may also tell or ask students to focus on the main topic or theme and to connect what they are reading to a larger organizing idea. For example, the teacher might stop reading after a few pages and ask "what have we read so far that gives us a clue about the moral of the story?"*

FAQs		
R16	***What if I only see one or two aspects of reading (e.g., just reading and post-reading) during the Observation Segment?***	*That's okay. Code only what you see during the observation, not what you think might have happened before or what might happen later.*
	How do I code if I observe an activity that is NOT whole group (i.e., small group or individual)?	*If the teacher works with small groups or individual students, follow the teacher as he or she moves from group to group or from student to student. Code what you observe with all groups/students the teacher worked with.*
	What if a teacher doesn't do all the talking, but asks students questions and the students do most or almost all the talking?	*This counts as talk and you should code these items.*

VI B. FOCUS ON MEANING OF BOOK/TEXT		Observation Segments					
DURING READING		**1**	**2**	**3**	**4**	**5**	**6**
R14 What did the teacher talk/ask about during reading? *(Code all that apply.)*	1 No reading observed or no talk during reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Talk relates to the book/text, but is not about the topic or content of the book.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Letters or words (sounding out letters or words; rhyming words; word recognition)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Vocabulary (word meaning)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Grammar/mechanics/spelling	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Key features of the book/text (type of book, parts of the book, author/illustrator)	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
	7 Text structure (parts of a story/text)	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐
	8 Reading comprehension strategies	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐
	9 The purpose for reading the text	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐
	10 Title, topic, subject and/or theme of text are reading	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐
	11 What happened in the story or what might happen next; OR what information was presented in the text	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐
	12 The characters in the text, who they are, their motivation and/or goals	12 ☐	12 ☐	12 ☐	12 ☐	12 ☐	12 ☐
	13 Connecting content with students' prior knowledge/experiences	13 ☐	13 ☐	13 ☐	13 ☐	13 ☐	13 ☐
	14 Other	14 ☐	14 ☐	14 ☐	14 ☐	14 ☐	14 ☐
R15 How much detail did the teacher use when talking about the content of the book/text during reading (#11, 12 or #13 from R14)? *(Code the highest.)*	1 No reading observed or no talk during reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No talk about content (did not code 11, 12 or 13 in R14)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Talk included 1-2 details about content	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Talk included 3 or more details about content	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
R16 How did the teacher organize the talk about content during reading? (#11, 12 or #13 from R14)? *(Code the highest.)*	1 No reading observed or no talk during reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No talk about content (did not code 11, 12 or 13 in R14)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Talked about content but only 1-2 details or details were not organized	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Talked about content and at least some of the details were organized	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐

FOCUS ON MEANING OF BOOK/TEXT : ITEMS R17 – R19

KEY WORDS AND PHRASES		
	Post-reading	*When the teacher closes the book or the reading activity has ended. Any type of activity, discussion, or teacher talk that directly follows a reading and that relates to the book/text is considered post-reading.*
	Announces end of reading activity	*This is a general announcement that signals that the reading activity is ending. For example the teacher might say "Ok, everyone put your books away; it's time for lunch."*
	Letters or words (sounding out letters or words; rhyming words; word recognition)	*Teachers may focus students' attention on letters or words.* *You may see two main approaches teachers use to help students recognize words: decoding and sight word recognition. First, decoding is when we sound out words that don't look familiar to us, to see if we recognize them. Teachers use many techniques to help develop students' abilities to sound out words, beginning with sounding out letters, as part of a reading activity or independently. Second, sight words are "high frequency" or common words: and, the, to, my, that, I, because, your, why, until, and first. Students are taught to recognize words by sight, rather than by sounding them out. Teachers will often practice sight words at morning meeting or during the sentence of the day.*
	Grammar/Mechanics/ Spelling	*Grammar refers to how we construct sentences, types of sentences and parts of speech, and all the rules involved (such as agreement, verb tenses). Mechanics include capitalization and punctuation. For example, a teacher may write a few sentences on the board for morning meeting, intentionally leaving out the periods at the end of the sentences, and then ask students questions about what needs to be fixed (add punctuation).*
	Key Features of the Text	*This includes the type of text it is (fiction/non-fiction; fable, adventure, science fiction, etc); and the authors and illustrators; the physical parts of the book: front cover, title page, beginning of text, end of text, back cover.*
R17	***Text Structure***	*Recognizing and using the author's organizational plan or text structure to help students understand and remember what they have read. Common elements of text structure include that each story has a beginning, middle and end, that there are main characters, characters' goals, key events, problems and solutions. Teachers often present a chart or diagram that shows the key aspects of text structure and/or have students complete a chart or diagram during or after reading. They might have students look up the sub-headings in a textbook chapter and create an outline of key points using the sub-headings.*
	Reading comprehension strategies	*Reading comprehension strategies involve the deliberate use of a cognitive routine by the reader before, during, or after reading. These cognitive routines are specific mental actions (such as previewing, summarizing, self-questioning, clarifying, and visualizing) that facilitate a better understanding of text.*
	Purpose of reading the text	*Teachers may explain, reiterate or elicit from students why the teacher had them read a particular text. For example, a teacher might say "Today we read this book to learn about dinosaurs."*
	Evaluating the text	*This includes when the teacher and/or students discuss whether the text was good, what about the text made it good or effective or successful.*
	Title, topic, subject and/or theme of text	*What the book or text is about, the story events or world information that one would find out by reading the text. This does not include decoding or naming single words.*
	What the text was about/ Summarizing	*This involves taking information from across the text to recap what happened in the story or the main points of an informational text.*
	The characters in the text, who they are, their motivation and/or goals	*This includes references to what the characters looked like, who they were, how they felt, what they wanted to accomplish, the reasons for their actions.*
R19	***Organization***	*Teachers may organize post-reading discussions by providing or eliciting a summary that organizes the information into categories or sequences of events ("Let's make a list of what we learned about George Washington's childhood, then about what he did during the Revolutionary War, then what he did as president, etc.").Teachers may ask students to retell parts or all of a story or to summarize the main points in a text they just read. These are also ways of organizing the post-reading talk.*

FAQs		
R17	***What if I only see one or two aspects of reading (e.g., just reading and post-reading) during the Observation Segment?***	*That's okay. Code only what you see during the observation, not what you think might have happened before or what might happen later.*
	How do I code if I observe an activity that is NOT whole group (i.e., small group or individual)?	*If the teacher works with small groups or individual students, follow the teacher as he or she moves from group to group or from student to student. Code what you observe with all groups/students the teacher worked with.*
	What if a teacher doesn't do all the talking, but asks students questions and the students do most or almost all the talking?	*This counts as talk and you should code these items.*

VI. B. FOCUS ON MEANING OF BOOK/TEXT		Observation Segments					
AFTER READING (POST READING)		**1**	**2**	**3**	**4**	**5**	**6**
R17 What did the teacher talk/ask about during post-reading? *(Code all that apply.)*	1 Did not observe end of reading or no post-reading talk	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Announces the end of the reading activity	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Letters or words (sounding out letters or words; rhyming words; word recognition)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Vocabulary (word meaning)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Grammar/mechanics/spelling	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Key features of the book/text (type of book, parts of the book, author/illustrator)	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
	7 Text structure (parts of a story/text)	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐
	8 Reading comprehension strategies	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐
	9 The purpose for reading the text	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐
	10 Evaluating the text	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐	10 ☐
	11 Title, topic, subject and/or theme of text read	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐	11 ☐
	12 What the text was about (summarizing)	12 ☐	12 ☐	12 ☐	12 ☐	12 ☐	12 ☐
	13 The characters in the text, who they are, their motivation and/or goals	13 ☐	13 ☐	13 ☐	13 ☐	13 ☐	13 ☐
	14 Connecting content with students' prior knowledge/experiences	14 ☐	14 ☐	14 ☐	14 ☐	14 ☐	14 ☐
	15 Other	15 ☐	15 ☐	15 ☐	15 ☐	15 ☐	15 ☐
R18 How much detail did the teacher use when talking about the content of the book/text during post-reading (#12, 13 or #14 from R17)? *(Code the highest.)*	1 Did not observe end of reading or no post-reading talk	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No talk about content (did not code 12, 13 or 14 in R17)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Talk included 1-2 details about content	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Talk included 3 or more details about content	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
R19 How did the teacher organize the talk about content during post-reading (#12, 13 or #14 from R17)? *(Code the highest.)*	1 Did not observe end of reading or no post-reading talk	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No talk about content (did not code 12, 13 or 14 in R17)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Talked about content but only 1-2 details or details were not organized	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Talked about content and at least some of the details were organized	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐

FOCUS ON MEANING OF BOOK/TEXT : ITEMS R20 – R21

KEY WORDS AND PHRASES

R20	***Connecting content with students' prior knowledge/experiences***	*This includes times when the teacher relates content from the book/text with students' prior knowledge and/or experiences. It includes times when the teacher invites students to make these connections as well as times when the students bring up the connection ("This character reminds me of my uncle.") and the teacher affirms and/or reinforces the connection ("Good job making connections."). Connecting prior knowledge to information in books/texts includes making connections from the student's personal experience, something the student learned previously either in or out of school, another area of study, or another text or book.*
	Big ideas	*Big ideas are main themes, concepts or lessons.*
R21	***Feedback***	*The response a teacher gives to a student's answer. We ask you to focus on several types:* *General feedback: Goes beyond simply letting students know that they were heard to affirming that they are doing a good job and/or that their efforts are appreciated and/or that they should keep on trying.* *Evaluative feedback: Lets students know when their answers are right or wrong (and perhaps why they are right or wrong).* *Specific feedback: When students struggle to respond to questions or with a task, teachers provide feedback that helps the students arrive at an answer or accomplish the task. OR If the students are not struggling, the teacher may give feedback by explaining how the students arrived at the answer.* *Strategic feedback: Whether the students are right or wrong, struggling or not, the teacher may ask students to explain their thinking.*

FAQs

R21	***What if I only see one or two aspects of reading (e.g., just reading and post-reading) during the Observation Segment?***	*That's okay. Code only what you see during the observation, not what you think might have happened before or what might happen later.*
	How do I code if I observe an activity that is NOT whole group (i.e., small group or individual)?	*If the teacher works with small groups or individual students, follow the teacher as he or she moves from group to group or from student to student. Code what you observe with all groups/students the teacher worked with.*
	What if a teacher doesn't do all the talking, but asks students questions and the students do most or almost all the talking?	*This counts as talk and you should code these items.*

VI. B. FOCUS ON MEANING OF BOOK/TEXT		Observation Segments					
BEFORE, DURING AND AFTER READING		**1**	**2**	**3**	**4**	**5**	**6**
R20 What parts of the content of the book/text did the teacher connect to students' prior knowledge/ experiences? *(Code all that apply.)*	1 No connections made	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Connections to the general topic of the book/text	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Connections to specific details in the book/text	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Connections to big ideas in the book/text	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
R21 When students answered questions about the content of the book/text, what type of feedback did the teacher provide? *(Code all that apply.)*	1 Students did not answer questions about content	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No feedback given or feedback was very vague (just confirmation that the teacher heard the student)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 General feedback (Good job!) or evaluative feedback (You did that right.)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Specific feedback that helps students arrive at an answer, complete a task, or teacher verbalizes how the student arrived at an answer.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Strategic feedback: asking students to explain how they figured out their answers or completed a task.	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐

DEFINING WORDS AS PART OF BOOK/TEXT SHARING: ITEMS R22-R27

KEY WORDS AND PHRASES		
R22-R27	***Definition***	*A definition explains, or shows the meaning of a word. Code any words defined during pre-reading, reading and/or post-reading, including words that are and are not from the book/text.*
R22 - R25	***Synonym***	*The teacher uses a different word with similar meaning to help explain the word (synonym). Example: "A mill is like a factory."*
	Antonym	*The teacher uses a word that means the opposite of the word she is trying to explain. Example: "The red dog was not tiny, it was huge."*
	Additional descriptors	*The teacher gives students additional details about what a word means, such as: "A mill is a place where people make flour. It is usually a tall building." To code this option, the teacher must have provided one of the other types of definitions. Using a word in a sentence also counts as an additional descriptor.*
	Pictures, visual representations, gestures, facial expressions, vocal quality	*Teachers may define what a word means by showing students a picture or drawing (such as pointing to a picture of a horse, or pointing to bicycles, cars, trains and planes to show what's meant by making gestures (moving arms and legs back and forth quickly to show running), facial expressions (sad face for unhappy), or changing their vocal quality (frightened voice to convey fear). Writing the word or definition does not count as a visual representation.*
R24 - R27	***Minimal involvement (quick answer, copying or repeating definitions)***	*This includes times when the teacher asks students if they agree or disagree with a definition, or direct students to repeat or copy a definition.*
	Some involvement (generating own definition)	*The teacher may ask students to speak or write their own definition for a word.*
	Extended involvement (classifying or comparing words; analyzing one word to explore its different meanings and uses; generating new examples)	*The teacher may ask students to do more than provide a word definition. The teacher may ask students to classify (these are all names for animals), compare (a cow eats grass and a chicken eats seeds), or to create charts, tables, graphs or diagrams that show different meanings a word might have. The teacher may ask students to generate new examples of how to use the word in a sentence.*

FAQs		
R22 - R27	***How do I code if I observe an activity that is NOT whole group (i.e., small group or individual)?***	*If the teacher works with small groups or individual students, follow the teacher as he or she moves from group to group or from student to student. Code what you observe with all groups/students the teacher worked with.*
	What should I code if a teacher gives the definition and ask the students to provide the word that matches?	*For the purposes of this study, we do not consider this "defining words," because the teacher is providing the definition and the students are basically providing the word or label.* *To be counted as "defining words" the teacher has to begin with the word and then either provide a definition (in any of the ways mentioned above) or solicit a definition of a word from students.*

VI. C. DEFINING WORDS AS PART OF BOOK/TEXT SHARING		Observation Segments					
BEFORE READING (PRE-READING)		**1**	**2**	**3**	**4**	**5**	**6**
R22 During pre-reading, how did the teacher and/or students define word(s)? *(Code all that apply.)*	1 Didn't observe beginning of reading or no talk before reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Did not define words before reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Provided a definition of a word (synonym/antonym or definition)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Provided additional descriptors/adjectives	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Showed a picture or visual representation of a word, or used a gesture, facial expression, or obvious vocal quality to convey word meaning.	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
R23 During pre-reading, did the teacher and/or students use more than one of the approaches listed in R22 to define <u>a single</u> word? *(Code only one.)*	1 Didn't observe beginning of reading or no talk before reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Did not define words before reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 No, used only one approach to define a single word	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Yes, used more than one approach to define a single word	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
R24 During pre-reading, did the teacher get the students involved in defining the word(s)? *(Code the <u>highest</u>.)*	1 Didn't observe beginning of reading or no talk before reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Did not define words before reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 No student involvement (only listening)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Minimal involvement (quick answer, copying or repeating definitions)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Some involvement (discussion or generating own definition)	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Extended involvement (classifying or comparing words; analyzing one word to explore its different meanings and uses; generating new examples)	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
DURING READING							
R25 During reading, how did the teacher and/or students define word(s)? *(Code all that apply.)*	1 No reading observed or no talk during reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Did not define words during reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Provided a definition of a word (synonym/antonym or definition)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Provided additional descriptors/adjectives	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Showed a picture or visual representation of a word, or used a gesture, facial expression, or obvious vocal quality to convey word meaning.	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
R26 During reading, did the teacher and/or students use more than one of the approaches listed in R25 to define <u>a single</u> word? *(Code only one.)*	1 No reading observed or no talk during reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Did not define words during reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 No, used only one approach to define a single word	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Yes, used more than one approach for a single word	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
R27 During reading, did the teacher get the students involved in defining the word(s)? *(Code the <u>highest</u>.)*	1 No reading observed or no talk during reading	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Did not define words during reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 No student involvement (only listening)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Minimal involvement (quick answer, copying or repeating definitions)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Some involvement (discussion or generating own definition)	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Extended involvement (classifying or comparing words; analyzing one word to explore its different meanings and uses; generating new examples)	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐

DEFINING WORDS AS PART OF BOOK/TEXT SHARING: ITEMS R28-31

KEY WORDS AND PHRASES		
R28-R30	***Definition***	*A definition defines, explains, or shows the meaning of a word. Code any words defined during pre-reading, reading and/or post-reading, including words that are and are not from the book/text.*
R28	***Synonym***	*The teacher uses a different word with similar meaning to help explain the word (synonym). Example: "A mill is like a factory."*
	Antonym	*The teacher uses a word that means the opposite of the word she is trying to explain. Example: "The red dog was not tiny, it was huge."*
	Additional descriptors	*The teacher gives students additional details about what a word means. They may give more details, such as: "A mill is a place where people make flour. It is usually a tall building." To code this option, the teacher must have provided one of the other types of definitions. Using a word in a sentence also counts as an additional descriptor.*
	Pictures, visual representations, gestures, facial expressions, vocal quality	*Teachers may define what a word means by showing students a picture or drawing (such as pointing to a picture of a horse, or pointing to bicycles, cars, trains and planes to show what's meant by making gestures (moving arms and legs back and forth quickly to show running), facial expressions (sad face for unhappy), or changing their vocal quality (frightened voice to convey fear). Writing the word or definition does not count as a visual representation.*
R30	***Minimal involvement (quick answer, copying or repeating definitions)***	*This includes times when the teacher asks students if they agree or disagree with a definition, or direct students to repeat or copy a definition.*
	Some involvement (generating own definition)	*The teacher may ask students to speak or write their own definition for a word.*
	Extended involvement (classifying or comparing words; analyzing one word to explore its different meanings and uses; generating new examples)	*The teacher may ask students to do more than provide a word definition. The teacher may ask students to classify (these are all names for animals), compare (a cow eats grass and a chicken eats seeds), or to create charts, tables, graphs or diagrams that show different meanings a word might have. The teacher may ask students to generate new examples of how to use the word in a sentence.*

FAQs		
R28 - R31	***How do I code if I observe an activity that is NOT whole group (i.e., small group or individual)?***	*If the teacher works with small groups or individual students, follow the teacher as he or she moves from group to group or from student to student. Code what you observe with all groups/students the teacher worked with.*
	What should I code if they give the definition and ask the students to provide the word that matches?	*For the purposes of this study, we do not consider this "defining words," because the teacher is providing the definition and the students are basically providing the word or label.* *To be counted as "defining words" the teacher has to begin with the word and then either provide a definition (in any of the ways mentioned above) or solicit a definition of a word from students.*

VI C. DEFINING WORDS AS PART OF BOOK/TEXT SHARING		Observation Segments					
AFTER READING (POST-READING)		**1**	**2**	**3**	**4**	**5**	**6**
R28 After reading, how did the teacher and/or students define word(s)? *(Code all that apply.)*	1 Did not observe end of reading or no post-reading talk	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Did not define words after reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Provided a definition of a word (synonym/antonym or definition)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Provided additional descriptors/adjectives	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Showed a picture or visual representation of a word or used a gesture, facial expression, or obvious vocal quality to convey word meaning.	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
R29 After reading, did the teacher and/or students use more than one of the approaches listed in R28 to define <u>a single</u> word? *(Code only one.)*	1 Did not observe end of reading or no post-reading talk	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Did not define words after reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 No, used only one approach to define a single word	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Yes, used more than one approach for a single word	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
R30 After reading, did the teacher get the students involved in defining the word(s)? *(Code the <u>highest</u>.)*	1 Did not observe end of reading or no post-reading talk	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Did not define words after reading	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 No student involvement (only listening)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Minimal involvement (quick answer, copying or repeating definitions)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Some involvement (discussion or generating own definition)	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Extended involvement (classifying or comparing words; analyzing one word to explore its different meanings and uses; generating new examples)	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐

R31 **List up to 10 words that were defined by the teacher and/or students in pre-reading, reading or post reading (coded in R22, R25, and/or R28).**

1 ☐ _____	5 ☐ _____	8 ☐ _____
2 ☐ _____	6 ☐ _____	9 ☐ _____
3 ☐ _____	7 ☐ _____	10 ☐ _____
4 ☐ _____		

DEFINING WORDS, NOT DURING BOOK/TEXT SHARING (VOCAB): ITEMS V1-V5

KEY WORDS AND PHRASES		
V1, V2	***Defining words, NOT during book/text sharing***	*In this dimension, code any time a teacher and/or student defines a word outside of a book/text sharing activity. Also, if, during pre-reading, reading and/or post-reading, any words are defined that are clearly NOT from the book/text, code them as VOCAB.*
V2	***Synonym***	*The teacher uses a different word with similar meaning to help explain the word (synonym). Example: "A mill is like a factory."*
	Antonym	*The teacher uses a word that means the opposite of the word she is trying to explain. Example: "The red dog was not tiny, it was huge."*
	Additional descriptors	*The teacher gives students additional details about what a word means. They may give more details, such as: "A mill is a place where people make flour. It is usually a tall building." To code this option, the teacher must have provided one of the other types of definitions. Using a word in a sentence also counts as an additional descriptor.*
	Pictures, visual representations, gestures, facial expressions, vocal quality	*Teachers may define what a word means by showing students a picture or drawing (such as pointing to a picture of a horse, or pointing to bicycles, cars, trains and planes to show what's meant by making gestures (moving arms and legs back and forth quickly to show running), facial expressions (sad face for unhappy), or changing their vocal quality (frightened voice to convey fear). Writing the word or definition does not count as a visual representation.*
V4	***Minimal involvement (quick answer, copying or repeating definitions)***	*This includes times when the teacher asks students if they agree or disagree with a definition, or direct students to repeat or copy a definition.*
	Some involvement (generating own definition)	*The teacher may ask students to speak or write their own definition for a word.*
	Extended involvement (classifying or comparing words; analyzing one word to explore its different meanings and uses; generating new examples)	*The teacher may ask students to do more than provide a word definition. The teacher may ask students to classify (these are all names for animals), compare (a cow eats grass and a chicken eats seeds), or to create charts, tables, graphs or diagrams that show different meanings a word might have. The teacher may ask students to generate new examples of how to use the word in a sentence.*

FAQs		
V1	***What if book/text sharing was the only activity I observed during the Observation Segment?***	*Check "1" for item V1, draw a line down the VOCAB items in that segment and proceed to the next dimension, COMP.* *VOCAB ONLY measures words defined during an activity OTHER than book/text sharing. (READ items R22-R31 measure words defined during book/text sharing.)*
V1-V4	***How do I code if I observe an activity that is NOT whole group (i.e., small group or individual)?***	*If the teacher works with small groups or individual students, follow the teacher as he or she moves from group to group or from student to student. Code what you observe with all of the groups/students the teacher worked with during the observation segment.*
V1-V5	***What should I code if they give the definition and ask the students to provide the word that matches?***	*For the purposes of this study, we do not consider this "defining words," because the teacher is providing the definition and the students are basically providing the word or label.* *To be counted as "defining words" the teacher has to begin with the word and then either provide a definition (in any of the ways mentioned above) or solicit a definition of a word from students.*

VII. DEFINING WORDS, NOT DURING BOOK/TEXT SHARING (VOCAB)		Observation Segments					
		1	2	3	4	5	6
V1 Did the observation segment include activities other than book/text sharing? *(Code only one.)*	1 No **(DRAW LINE DOWN SEGMENT & GO TO COMP)**	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Yes **(CONTINUE CODING THIS DIMENSION)**	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
V2 How did the teacher and/or students define word(s)? *(Code all that apply.)*	1 No words defined	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Provided a definition of a word (synonym/antonym or definition)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Provided additional descriptors/adjectives	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Showed pictures or visual representations of a word, or used gestures, facial expressions, or obvious vocal quality to convey word meaning.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
V3 Did the teacher and/or students use more than one of the approaches in V2 to define <u>a single</u> word? *(Code only one.)*	1 No words defined	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No, used only one approach to define a single word	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Yes, used more than one approach to define a single word	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
V4 Did the teacher get the students involved in defining the word(s)? *(Code all that apply.)*	1 No words defined	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 No student involvement (only listening)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Minimal involvement (quick answer, copying or repeating definitions)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Some involvement (discussion or generating own definition)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Extended involvement (classifying or comparing words; analyzing one word to explore its different meanings and uses; generating new examples)	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐

V5	List up to 10 words that were defined by the teacher and/or students, NOT during book/text sharing.		
	1 ☐ _____	5 ☐ _____	8 ☐ _____
	2 ☐ _____	6 ☐ _____	9 ☐ _____
	3 ☐ _____	7 ☐ _____	10 ☐ _____
	4 ☐ _____		

READING COMPREHENSION STRATEGIES/SKILLS (COMP): ITEMS C1-C5

KEY WORDS AND PHRASES		
C1 Identify and use or discuss reading comprehension strategies		*To code for COMP, the teacher has to identify a strategy/skill, by either naming or explaining the strategy/skill PLUS use the strategy/skill OR discuss its use.*
C1-C5 Types of Reading Comprehension Strategies and Skills	***Previewing***	*Going through book/text before reading, without reading the book/text, to get a sense of its content or structure (e.g., what it's about). This is often called a picture walk.*
	Predicting	*Making guesses about what might happen next in a book/text.*
	Connecting to prior knowledge	*Helping students relate what they already know to what is being read in order to better understand the meaning of the text. The teacher may refer to this as "making connections."*
	Summarizing	*Briefly describing the main points or main content of the book/text verbally or in writing.*
	Visualizing/Sensory Imaging	*Creating a mental picture of the story based on the language and other clues in the text (not simply describing a picture or illustration). Asking students to imagine what something smells, sounds, feels or tastes like.*
	Text Structure	*Recognizing and using the way that authors organize information in text in order to better understand the meaning of the text. This includes calling attention to the headings and sub-headings in non-fiction texts, and to aspects of story structure (setting, characters, beginning, middle, end, problem/solution).*
	Questioning/ reacting to text while reading	*Teacher may demonstrate how students can ask themselves questions and/or react to text while reading to enhance their understanding of the text. Or, they may have students practice the questioning/reacting to text.*
	Self-Monitoring	*Helping students self-assess whether they are understanding what they are reading. For example after every few pages the teacher asks students to stop, ask themselves if they understood what they read (i.e. could they summarize what they read in their own words) and if not, re-read the passage.*
	Inferences/Drawing Conclusions	*Using information or clues from the book/text to identify character motivations and emotions or the big idea or main theme of the book/text. The teacher may refer to this as "making connections."*
C3 General explanation of how		*The teacher explains, in general, how to use the strategy. ("Let's make predictions and make guesses about what we think the book will be about.")*
Specific directions or demonstrating how to use a strategy		*The teacher explains how to use the strategy, including breaking down the strategy, step by step. For example:* *"When we predict, we look at the cover and ask, What do we think is going to happen in this story? Then after every few pages, we ask, Do I need to change my predictions based on what I just read?"*
Demonstrating		*The teacher shows students how to implement the strategy by using it him or herself, while the students watch, and talking about what he or she is doing.*
C5 Using strategies for specific types of text		*The teacher may explain that particular reading comprehension strategies are useful when reading specific types of text. For example, a teacher may note that predictions are helpful when reading a mystery book or that using prior knowledge is important when reading a science text.*

FAQs		
C1	***What if none of the types of reading comprehension strategies are identified or used during the Observation Segment?***	*Code "1- No Strategy" for item C1, draw a line down the COMP items for that segment, and proceed to the next dimension, KNOW.*
	What if the teacher only names a strategy, but does not discuss it, use it or have students use it?	*Code "1- No Strategy" for item C1, draw a line down the COMP items for that segment, and proceed to the next dimension, KNOW.*
	What if a teacher describes a strategy but does not label it, or does not label it with the term we use to define that strategy?	*As long as the teacher is describing a strategy that is cited above that counts, even if he or she does not label it with the term we used above.*
	What if a student uses a strategy independently/spontaneously? (e.g., not prompted by teacher to do so).	*If a student independently or spontaneously uses a reading comprehension strategy, the teacher has to label, discuss or use that strategy in order to code a 2 or a 3 for C1. This is because this item measures what strategies teachers teach, not what strategies students use.*
	If the teacher labeled the student's independent use and then asked that student or another student to practice the strategy again, does that count?	*If the students are using a strategy, and the teacher calls attention to it by labeling it (i.e., "I like how you revised your predictions."), then code for COMP.*
	Does the teacher need to use or simply ask the students to use the strategy?	*Both count. If the teacher labels the strategy and asks the students to use it, code for COMP. If the teacher labels the strategy and uses it, code for COMP.*
	Can a comprehension strategy be used during pre-reading and/or post-reading, or does it have to be used during reading?	*A comprehension strategy can be used at any point during the reading session: pre-reading, during reading, and/or post-reading. For example, a teacher may have students generate predictions before they read, have them revise their predictions while they read, and then revisit them a final time after they've read a text.*
C4	***What if the teacher shows them WHEN to use the strategy but never explains or discusses WHEN?***	*Then code a "1" for C4. The teacher needs to explicitly explain WHEN, not just show WHEN a strategy is used.*

VIII. READING COMPREHENSION STRATEGIES/SKILLS (COMP)		Observation Segments					
		1	2	3	4	5	6
C1 Did the teacher identify AND use or discuss reading comprehension strategies/skills? *(Code only one.)*	1 No strategy **(DRAW LINE DOWN SEGMENT & GO TO KNOW)**	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 One strategy	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 More than one strategy	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
C2 Did the teacher explain WHY at least one strategy/skill should be used? *(Code the highest.)*	1 No explanation of why	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 General affirmation of strategy (Good readers do this.)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 General explanation of the purpose of strategy (to help us understand what we read; to help us remember)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Specific explanation of the purpose (to help us remember what we read by giving us a way to organize the information in the text)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
C3 Did the teacher explain or demonstrate HOW to use at least one strategy/skill? *(Code the highest.)*	1 No explanation of how (or explanation may be unclear)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 General explanation of how	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 **Specific directions** on how to use the strategy, (such as breaking down the strategy into steps) or **demonstrating** how to use the strategy	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
C4 Did the teacher explain WHEN to use at least one strategy/skill? *(Code the highest.)*	1 No explanation of when	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Explains when the strategy can be used (before, during, and/or after reading)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
C5 Did the teacher state that the strategy/skill is used when reading a specific type of text? *(Code only one.)*	1 No	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Yes	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐

READING COMPREHENSION STRATEGIES/SKILLS (COMP): ITEMS C6-C7

KEY WORDS AND PHRASES		
C6	***Specific guidance***	*This includes asking students questions, giving them reminders, and providing them with visual aids such as diagrams or outlines to help them use the strategy.* *It also includes explaining how to use the strategy including breaking down the strategy, step by step.*
	Parts of the reading activity	*This includes pre-reading, during reading, and post-reading activities.*
C7	***Feedback***	*The response a teacher gives to a student's answer. We ask you to focus on several types:* *General feedback: Goes beyond simply letting students know that they were heard by affirming that they are doing a good job and/or that their efforts are appreciated and/or that they should keep on trying.* *Evaluative feedback: Letting students know when their answers are right or wrong (and perhaps why they are right or wrong).* *Specific feedback: When students struggle to respond to questions or with a task, teachers provide feedback that helps the students arrive at an answer or accomplish the task. OR If the students are not struggling, the teacher may give feedback by explaining how the students arrived at the answer.* *Strategic feedback: Whether the students are right or wrong, struggling or not, the teacher may ask students to explain their thinking.*

FAQs		
C6	***What if the teacher helps one student one time and another student one time? Does this count as two times?***	*No. If the teacher helps multiple students once each, then COMP6 is coded as 3 – provided specific guidance during one part of the reading activity. The teacher needs to help a single student multiple times to code COMP6 a 4.*
	What if the student uses a strategy and talks about it, does this count?	*If the student uses a strategy (such as making a prior knowledge connection) and the teacher calls attention to it, either by naming it or explaining it, then it counts.*
	If the teacher is asking questions during the picture walk or otherwise involving the students, does that count as the students using a strategy?	*Yes, if the teacher involves the students in the preview or picture walk activity (i.e. by directing them to look at the pictures or asking them questions) that counts as the students practicing the strategy.*
C7	***What if the teacher asks one or several students to use a strategy in front of the class but does not provide feedback?***	*Code COMP 7 as "2" Students used strategies/skills, but no teacher feedback was provided.*
	What if a teacher says "OK" when students answer a question?	*This counts as "vague feedback" and would be coded as a "2." OK means "I hear you," but doesn't tell students whether their answers were right or wrong, on-target or off-target.*

VIII. READING COMPREHENSION STRATEGIES/SKILLS (COMP) (continued)		Observation Segments					
		1	2	3	4	5	6
C6 When students used comprehension strategies, did the teacher provide specific guidance about how to use the strategies/skills? ***(Code the highest.)***	1 The teacher did not have students use the strategies/skills.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 The teacher had the students use the strategy, but provided no guidance.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 The teacher provided specific guidance during one part of the reading activity.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 The teacher provided specific guidance during more than one part of the reading activity.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
C7 When students used comprehension strategies, did the teacher provide feedback? ***(Code the highest.)***	1 Teacher did not have students use strategies/skills.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Teacher had students used strategies/skills, but provided no feedback or feedback was very vague (just confirmation that the teacher heard the student)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Teacher gave general encouragement (Good job!) or evaluative feedback (You did that right!).	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Teacher gave specific feedback on how, when and/or why to use the strategy/skill better (Remember to revise your predictions in the middle of the story.)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Teacher gave strategic feedback, asking students to explain how, when and/or why to use the strategy/skill (Why did you underline the sub-headings in the chapter on climate? Tell me the process you used to revise your predictions while you were reading.).	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐

WORLD KNOWLEDGE (KNOW): ITEMS K1-K8

KEY WORDS AND PHRASES		
K1-K9	***World Knowledge***	*World knowledge includes general knowledge and content knowledge. It does not include skills or strategies (e.g., reading comprehension strategies, math computational skills).* ***General Knowledge*** *includes information about how and why things work (such as time, weather), how people live (such as where they work, what they eat, how they play), how people live together (families, communities, countries). In younger grades you will see general knowledge taught/reviewed during morning meeting such as the days of the week, the months of the year, etc. General knowledge serves as the foundation for content knowledge.* ***Content Knowledge*** *includes subject-specific facts and concepts, such as facts or concepts in geography, history (such as historical figures) civics, health, science (life cycles, planets, moon, stars), mathematics, and the arts. Content knowledge builds upon general knowledge. Note that, for mathematics, computational skills are not content knowledge. Understanding concepts (such as the different shapes and recognizing patterns) is content knowledge.*
K1 ***K2***	***Introduce or reinforce***	*Teachers often introduce ideas/concepts/information about a world knowledge topic through explanation or example. In the middle of a reading lesson, they might reinforce world knowledge by asking questions that call attention to the world knowledge in the text/book or questions that link something students have learned to the content they are reading.*
K4	***Literary concepts***	*Includes the different types of literature (novels, plays, poems) and genres (fiction, non-fiction, biography, mystery, romance, science fiction, etc). It also includes references to literary techniques, such as symbolism, metaphor and imagery. It does not include references to the author, illustrator, or publisher.*
K5	***Naming things***	*This includes the teacher naming and/or teacher having students name things -- objects, places, events, actions, people. Note that creating lists of names such as presidents names, names of animals, names of activities) are coded as "naming" not as "facts".*
	Reviewing and/or discussing facts	*Reviewing facts involves going beyond simply naming things to discussing information about things. For example, students may list all of the animals that they have as pets (a "naming things" activity) and then note one important fact about each animal (a "reviewing facts" activity).*
	Providing a definition of a word or concept	*This includes times when the teacher defines a word or concept related to world knowledge (such as defining "life cycle" in a science lesson, or defining "freedom" when reading a fictional story about colonial times). It also includes times when the teacher asks students to define words or concepts.*
	Presenting detailed information about a topic	*This refers to times when the teacher presents detailed information, rather than engaging students in a brief review or discussion of information already learned. The topic means the subject at hand – what's being discussed or studied; can be a person, place, event , object, or concept (such as George Washington, mountains, Thanksgiving, the Civil War, bridges, freedom). This also includes times when students are giving presentations.*
	Using technology or multi-media	*Includes using computers, internet, smart boards, MP3 players and any other technology, as well as showing videos or Power Point presentations, or playing songs with words.*
	Hands-on activities	*Includes using scientific or mathematical tools (such as rulers, scales and/or other tools to measure things), conducting experiments, building structures, creating visual representations of ideas, acting out what they have learned.*
	Other	*This includes activities such as having students interview family members or other students and activities that involve writing.*
K6	***Big ideas or themes***	*Big ideas or themes are recurring ideas or concepts that teachers use to enhance students' learning. For example, "freedom means responsibility" might be a theme for a 3rd grade class. The teacher would have students read history books about the revolutionary war and fictional books about colonial times, and interview their family about what freedom means to them. Thus, the theme helps students relate what they learn in social studies and reading/language arts. Topics are like the pieces of a puzzle; bid ideas/themes are the way teachers pull the puzzle together.*
K8	***Different pieces of information***	*With this question, we capture information about the breadth of information that students are exposed to during the segment. For example, in a lesson about George Washington, do the students hear a few facts (less than 5 pieces of information), some facts (5-10 pieces of information) or many facts (more than 10 pieces of information)? Pieces of information are individual facts about a topic.*

FAQs		
K1	***What if no world knowledge is covered during the Observation Segment?***	*Check "1"– "No world knowledge Covered" for item K1, draw a line down the KNOW items for that segment, and proceed to the next dimension, HIGH.*
K1, K5	***When students work in centers, playing at being police officers or other professions, does this count as world knowledge.***	*When following the teacher, if he or she joins a group of students who are playing at professions, you would count their play as building world knowledge. When following the teacher during center time, be careful to look for times when students are practicing or applying world knowledge, such as pretending to have jobs, building towns, doing experiments.*
	During the segment, I observed the last minute of a social studies lesson, where the teacher defined the word "democracy. I coded this under VOCAB. Do I still code it for KNOW as well?	*Yes, although it was brief and although you coded it for VOCAB, you code this type of event for KNOW as well.*
K2	***What if some students leave or new students join the group during the lesson? How do I code item K2?***	*You should count the total number of students exposed to the content, even if students are in small groups, working individually, or if some students entered or left the group during the lesson. That is, you should still count students who left before the lesson was over or students who joined the group part way through the lesson.*
K1, K4	***I'm observing a Pre-K or K classroom and am having trouble identifying world knowledge.***	*Pre-K and K classrooms are more likely to focus on developing students' general knowledge, although some content knowledge may also be taught. Pay special attention to center-based activities, and see the manual for specific examples of activities and concepts that fulfill the requirements of KNOW.* *Grades 1 through 3 are more likely to focus on building students' content knowledge in subjects like science, social studies and mathematics, although general knowledge may also be taught at any time.*
K3	***What if the observation segment is less than 15 minutes? If the segment is only 10 minutes long and the whole time the teacher introduces world knowledge, how do I code K3?***	*You would code it All or most of the observation segment (10-15 minutes).*
K5	***What is the difference between reviewing or discussing facts (2) and presenting detailed information (4)?***	*For 2, the students are involved in some way in reviewing or discussing information. For 4, the teacher presents the information. After reading a book about animals, if the teacher asks the students to list the ways that frogs and fish are alike and different, they are discussing facts (2). If he or she presents a detailed explanation of all the ways that frogs and fish are alike and different, this is a presentation (3).*

IX. WORLD KNOWLEDGE (KNOW)		Observation Segments					
		1	2	3	4	5	6
K1 Did the teacher introduce, reinforce or otherwise teach world knowledge? *(Code only one.)*	1 No world knowledge covered **(DRAW LINE DOWN SEGMENT & GO TO HIGH)**	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Yes **(CONTINUE CODING THIS SECTION)**	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
K2 To how many students in the class did the teacher introduce, reinforce or teach world knowledge? *(Code only one.)*	1 Two or fewer students	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Less than half the class	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Half the class or more	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 The whole class	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
K3 For how much time did the teacher introduce, reinforce or teach world knowledge? *(Code only one.)*	1 Very briefly (less than 5 minutes)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Some of the observation segment (5-10 minutes)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 All or most of the observation segment (10-15 minutes)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
K4 What content was the world knowledge related to? *(Code all that apply.)*	1 Social Studies (details about real people, jobs, types of food, current events, history, geography, government, money, the arts, religion, language, and famous people)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Health and Science (including animals, weather, nutrition, states of matter, life sciences, and scientific method)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Math (patterns, measurement, shapes, time, days of the week, calendar)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Literary concepts (types of literature, symbolism, metaphor, imagery)	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Other	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
K5 What approach(es) did the teacher use to introduce, reinforce or teach world knowledge? *(Code all that apply.)*	1 Teacher and/or students named or listed things (objects, places, events, actions, people).	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Teacher and/or students reviewed or discussed facts.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Teacher and/or students provided a definition of a word or concept.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Teacher presented detailed information about a topic (or topics).	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐
	5 Teacher read to students about a topic.	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐	5 ☐
	6 Teacher had students read about a topic.	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐	6 ☐
	7 Teacher and/or students used technology or multimedia.	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐	7 ☐
	8 Teacher had students engaged in hands-on activities.	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐	8 ☐
	9 Other	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐	9 ☐
K6 Did the teacher relate the information about world knowledge to a big idea or theme? *(Code only one.)*	1 No	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Yes	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
K7 Did the teacher actively involve students in learning world knowledge? *(Code only one.)*	1 No, students listened to the teacher or read silently.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Yes, students read out loud, discussed or answered questions, wrote, drew, acted, sang.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
K8 How many different pieces of information about world knowledge did the teacher and/or students talk about? *(Code only one.)*	1 Less than 5 pieces of information	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 5-10 pieces of information	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 More than 10 pieces of information	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐

WORLD KNOWLEDGE (KNOW): ITEM K9

KEY WORDS AND PHRASES		
K9 Connecting information about the world to students' prior knowledge		*This includes times when the teacher relates world knowledge with students' prior knowledge and/or experiences. It includes times when the teacher invites students to make these connections as well as times when the students bring up the connection ("My uncle is a fireman, like the man in this book.") and the teacher affirms and/or reinforces the connection ("Good job making connections."). Connecting prior knowledge to information includes making connections from the student's personal experience, something the student learned previously either in or out of school, another area of study, or another text or book.*
K9 Types of connections to prior knowledge	***Student's personal experiences***	*This connection is between the content and something the student has experienced. These can be emotional experiences (feeling angry, sad), family experiences (holidays sibling rivalry), or physical experiences (having a haircut, losing a tooth).*
	Connections to something previously learned, in any content area (subject in school) or context (out of school)	*The teacher links world knowledge to something learned formally, either in school or outside of school, in any content area – mathematics, social studies, science, the arts. ("Today we are going to read a book about animals. Remember when we studied the types of animals that live at the zoo?")*
	A previously read text/book.	*You should select this type of connection when you hear a teacher or student establish a clear link between the current lesson and a specific book or a text they have read in the past. ("Today we are going to read another book about animals. Last week we read the book about animals that live at the zoo. I wonder where the animals in this book are going to live").*

FAQs		
K9	***To count as a prior knowledge connection, does the teacher have to check to make sure students understand the prior knowledge connection?***	*No. As long as the teacher makes an attempt to connect world knowledge to students' prior knowledge, it counts. The teacher does not have to follow up to make sure the students were able to make this connection.*
	What if the teacher refers to a book that they just finished reading? How should I code K9?	*It does not matter how long ago the teacher or students read the book/text. If the teacher makes a connection between the current activity and a book students read previously, count it as a prior knowledge connection.* *Note: If the activity involves reading a text/book, then referring to this text/book doesn't count as a prior knowledge connection.*

IX. KNOWLEDGE OF THE WORLD (KNOW) (continued)			Observation Segments					
			1	2	3	4	5	6
K9	**How did the teacher connect information about the world to students' prior knowledge?** ***(Code all that apply.)***	1 No connections made with prior knowledge	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
		2 Connections to students' personal experiences	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
		3 Connections to something previously learned, in any content area (subject in school) or context (out of school)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
		4 Connections to a book/text previously read	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐

HIGHER-ORDER THINKING ITEMS H1-H5

KEY WORDS AND PHRASES		
H1 Encourage higher-order thinking		*Teachers encourage students to think beyond basic facts by using a variety of techniques. The most common is through asking questions that involve analysis, synthesis application of knowledge, evaluation, creative thinking or explaining thinking. They also may engage students in writing tasks that involve higher-order thinking and in hands-on tasks, like experiments, creative tasks (like creating a play or skit, writing a song, and illustrating stages in a plant's growth cycle).*
H1-H3 Higher- Order Thinking	***Application of Knowledge***	*Solving problems in new situations by applying knowledge, facts, and rules, such as correctly identifying a bear as a mammal or conducting an experiment.* *Pre-K – Grade 3: The teacher reads a fable to the class and asks each student to state what they think the moral of the story is. She then asks them to say or write what they would have done if they were the main characters.*
	Analysis	*Exploring relationships among information and ideas, such as classifying, comparing/contrasting, examining cause and effect, problem and solution, sequencing and/or ordering ideas or information* *Pre-K: Students plant a seed and draw pictures every week of the growing flower. They then make a book and present the book to their parents, explaining the changes in the flower, from seed to bloom.* *Grade 2: Students plant a seed and draw pictures every week of the growing flower. They make a book, writing at the bottom of each picture a description of the changes they see and on the last page labeling the parts of the flower.*
	Evaluation	*Making judgments about information or ideas based on a set of criteria, such as rating or ranking.* *Pre-K/K: Teacher holds up a mixture of drawings of children behaving well and misbehaving. She asks the students to identify which students are following the class rules and which are not.* *Grades 1-3: Students work in partners to read each other's writing and check to see if their partner fully responded to the question.*
	Synthesis	*Combining information or ideas in order to draw conclusions or make inferences.* *Pre-K/K: The teacher asks students to look at the picture in the book and say why the main character is laughing. The students have to consider what is happening on the page and make an inference about the main character's emotions, based on what is happening at that part of the story (represented in the pictures).* *Grades 1-3: The teacher asks students to decide what the moral of a fable is. The students have to think about the whole story and draw a conclusion about what the main lesson is.*
	Creative Thinking	*Developing new ideas, new solutions, and/or new approaches, such as finding new ways of performing tasks or new ways to learn – "thinking outside the box." (NOTE: The focus here is on creative thinking, not on creative arts.)* *Pre-K-Grade 3: The teacher reads a story to the students and asks each student to come up with a new ending for the story.* *Pre-K-Grade 3: The teacher writes a sentence starter on the board: "Once upon a time, a boy found a pair of magic shoes." She reads the sentence out loud and goes around the room, asking each student in turn to continue the story, so that the whole class has created a story together.*
	Explanations of Thinking	*Asking students to explain how they arrived at an answer or conclusion such as asking students to think out loud when solving a problem and/or performing a task.* *Pre-K: A student is busy building in the blocks center. The teacher asks the student to explain why she selected the types of blocks she's using.* *Grade 2: The teachers had the students read a paragraph that was very hard. They were allowed to use references and to consult with each other to help them understand the text. She then asked them to explain what strategies they used to try to understand the text.* *Grade 3: The students complete a brief science experiment that involves drawing a conclusion. The teacher asks them to explain how they arrived at their conclusion.*
H3 Examples of higher-order thinking questions		*How are these two cars the same? How are they different? If I drop an ice cube in this full cup of water, what will happen?* *What is the moral of the story we just read? How do you know? Which picture looks more like the weather we're having today?* *Why do you think so? If you were writing this story, how would you have it end?*
H5 Explaining answer and/or thinking		*When students have answered higher-order thinking questions, teachers may ask them to explain their answer or their thinking as a way of reinforcing higher-order thinking and students' awareness of thought processes.*

	FAQs	
H1	***What if the teacher did not encourage any higher-order thinking during the segment?***	*Check "1"– No Higher-Order Thinking" for item H1, draw a line down the HIGH items in that segment, and proceed to the next dimension, SUMM.*
H3, H5	***How do I know if a question that the teacher asked encouraged higher-order thinking?***	*If the question that the teacher asks requires students to do more than report a single fact or list of facts, then higher-order thinking was encouraged. Often teachers do this by asking 'how' or 'why' questions. For example: "When was Thomas Jefferson born?" Is a fact question. "Why do people think Thomas Jefferson was a good president?" is a higher-order question.*
H3	***What if the teacher asks the same higher-order question several times? Does it count as one question or multiple questions?***	*It counts as one question. This item isn't about the number of questions asked, but about the number of different higher-order questions the teacher asked.*
H3-H5	***What if the activity doesn't include any questions?***	*During the segment, you might observe students reading or writing, but not answering questions. If you can observe what the assignment is, then code it as a question. If you cannot, then code a "1" (no higher-order questions).*
H5	***How long should teachers wait for a response after they ask a student a question?***	*The teacher should wait at least 3-5 seconds for a student response.*
	What if the teacher only asks one high-order question and waits for 5 seconds for the student to answer?	*Code this as "4" for always allowed, even though it's based on only one question.*
	What if students call out answers as soon as the teacher asks the question, and the teacher does not ask them to wait?	*Code this as "2" to indicate that the teacher did not have a routine in place to allow all students time to think. This may happen frequently in younger grades.*

X. HIGHER-ORDER THINKING		Observation Segments					
		1	2	3	4	5	6
H1 **Did the teacher encourage students to use higher-order thinking?** ***(Code only one.)***	1 No encouragement of higher-order thinking ***(DRAW LINE DOWN SEGMENT & GO TO NEXT OBSERVATION SEGMENT OR TO SUMM)***	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Yes, teacher encouraged higher-order thinking. **(CONTINUE CODING THIS SECTION)**	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
H2 **For how much time did the teacher encourage students to use higher-order thinking?** ***(Code only one.)***	1 Very briefly (less than 5 minutes)	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Some of the observation segment (5-10 minutes)	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 All or most of the observation segment (10-15 minutes)	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
H3 **How many different questions did the teacher ask that encouraged students to use higher-order thinking?**	**RECORD #**						
H4 **Of the higher-order questions, how many were questions that asked students to explain their answer(s) or thinking?**	**RECORD #**						
H5 **Did the teacher allow time (3-5 seconds) for students to respond to the higher-order thinking questions?** ***(Code only one.)***	1 Teacher didn't ask higher-order thinking questions.	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐	1 ☐
	2 Teacher didn't allow time for students to respond to higher-order thinking questions.	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐	2 ☐
	3 Teacher sometimes allowed time for students to respond to higher-order thinking questions.	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐	3 ☐
	4 Teacher almost always allowed time for students to respond to higher-order thinking questions.	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐	4 ☐

SUMMARY: CLASSROOM CLIMATE (SUMM):ITEMS SM1 – SM5

KEY WORDS AND PHRASES		
SM1	***Teacher***	*Remember that by "teacher" we mean any adults you observed leading instruction during your observation period (across all segments).*
	Praised students' school work	*This includes praise in response to students' school work, to what children say, do and write.*
	Responded with interest to students' comments or activities	*This includes responding to children personally, such as greeting children in the morning and asking about their weekend, commenting on things the children like or don't like, and otherwise showing a personal interest in the students.*
	During reading, clearly conveyed warmth	*Code this option if, during book/text sharing, the teacher used a warm and/or encouraging tone of voice, facial expressions and/or gestures. Code this if, the teacher created a warm atmosphere, that likely made students feel at ease during the reading.*
SM3	***Physically harsh***	*This includes any physical action that attempts to force a child to do something, such as pushing a child so they get into line, pulling at their bags, or using physical size to intimidate a child.* *This does not include gentle touching such as touching a child on the shoulder to get his or her attention, ruffling a child's hair to encourage him or her to stand in line.*
	Verbally harsh	*This includes excessive use of volume to reprimand students (screaming or yelling at individual children, groups or the whole class), but not yelling to get the class' attention. Includes criticizing individual children, groups or the whole class, rather than correcting, especially any personal criticisms.*
	Left students unsupervised	*Students are alone with no other adult present. If the teacher steps outside the room to speak to an adult or a child but can still see the children, then they are still supervised. If the teacher leaves the students unsupervised for any amount of time code this.*
	Deliberately ignored student's question/ comment or request for help	*This includes when a teacher disregards a student's request for help. This does not include when a teacher fails to call on all the students who have their hands raised (unless it is obvious that the teacher is intentionally and unfairly ignoring a particular student and is calling on other students).The teacher/adult should respond to students' questions with some indication that he or she can see or hear them.*
	Was sarcastic or embarrassed students	*The teacher uses humor or sarcasm that makes fun of a student or calls attention to something embarrassing.*
	Placed frequent restrictions on students	*This includes frequently restricting children's talk (such as no talking except talk directed by the teacher) or movement (such as requiring that children sit still and at attention at all times).*
	Displayed favoritism	*The teacher may show one or two students in the class more attention than other students. The other students may feel that they are less important or able to learn than the favored students. Code this item if the teacher displayed favoritism toward some students during the observation.*
	Ignored students' physically aggressive behavior	*Two students are getting into a fist fight and the teacher ignores them and does not stop them.*
	Ignored students' verbally aggressive behavior	*A student makes fun of another; the class laughs, and the teacher ignores this.*

FAQs		
SM1-SM5	***When should I code these items?***	*Unlike the other items in the rubric, SUMM items should be coded at the end of a full observation, as they measure the classroom climate across the observations segments. Please take into account what you observed during all segments. You record the types and frequency of positive and negative interactions so it's easy to total them up at the end of the full observation.*
SM2	***When coding how many positive interactions the teacher had with students, are we supposed to be considering the experience of any one student or the experience of the majority of students?***	*Tabulate how many positive interactions the teacher had with any student, across all the observation segments in your observation session.*
SM5	***When coding how many students had negative interactions with the teacher, do I count students who heard or saw the negative interaction or only the person directly affected?***	*Only count those students to whom the negative interaction was directed.*

XI. SUMMARY: CLASSROOM CLIMATE (SUMM)			Only code after last segment
SM1 What types of positive interactions did the teacher have with students? ***(Code all that apply.)***	1	No instances of positive interactions	1 ☐
	2	Smiled at or laughed with students	2 ☐
	3	Respectfully listened to students	3 ☐
	4	Used a warm, calm voice	4 ☐
	5	Praised or commented positively on students' school work	5 ☐
	6	Responded with interest to students' comments or activities	6 ☐
	7	Drew attention to positive child behavior	7 ☐
	8	Used nonverbal responses to student (nod, wink, talking to students at eye level, thumbs up, high fives)	8 ☐
	9	During book/text sharing, clearly conveyed warmth	9 ☐
SM2 How many positive interactions did the teacher have with students? ***(Code only one.)***	1	None	1 ☐
	2	A few positive interactions	2 ☐
	3	Multiple positive interactions	3 ☐
	4	Consistent positive interactions	4 ☐
SM3 What types of negative interactions did the teacher use/allow? ***(Code all that apply.)***	1	No instances of negative interactions	1 ☐
	2	Was physically harsh	2 ☐
	3	Was verbally harsh	3 ☐
	4	Left students unsupervised	4 ☐
	5	Deliberately ignored student's question/comment or request for help	5 ☐
	6	Was sarcastic or embarrassed students	6 ☐
	7	Placed frequent restrictions on students	7 ☐
	8	Displayed favoritism	8 ☐
	9	Ignored student's physically aggressive behavior	9 ☐
	10	Ignored student's verbally aggressive behavior	10 ☐
SM4 How many negative interactions did the teacher have with students? ***(Code only one.)***	1	None	1 ☐
	2	One or two negative interactions	2 ☐
	3	Three or four negative interactions	3 ☐
	4	Five or more negative interactions	4 ☐
SM5 How many students had negative interactions with the teacher? ***(Code only one.)***	1	None	1 ☐
	2	A few students	2 ☐
	3	Many students	3 ☐
	4	Whole class	4 ☐

SUMMARY: CLASSROOM CLIMATE (SUMM): ITEMS SM6-SM7

KEY WORDS AND PHRASES		
SM6	***Lead teacher***	*This is the main teacher. The lead teacher is the person who is responsible for the classroom and designated the lead teacher. If there are co-teachers, where both teachers assume full responsibility for the class, please code them both as lead teachers.*
SM7	***Other adult***	*Any other adult you coded for during the observation segments. This could be a co-teacher, assistant teacher, aide, specialist, another teacher, a parent or a visitor.*

FAQs		
SM6	***How do I fill out SM6 if there are co-teachers?***	*Generally you will only code one for this item however you can code more than one option in SM6 if there are multiple lead teachers (aka co-teachers). Fill out one language option for each lead teacher.*

X. SUMMARY: CLASSROOM CLIMATE (SUMM) (continued)		Only code after last segment
SM6 Language(s) spoken by LEAD TEACHER ***(Code only one unless co-teachers. Code all that apply for co-teachers.)***	1 English only	1 ☐
	2 Spanish or other language only	2 ☐
	3 Primarily English, some Spanish or other language	3 ☐
	4 Primarily Spanish or other language, some English	4 ☐
	5 English and Spanish (or other) about equally	5 ☐
SM7 Language(s) spoken by other adults observed ***(Code all that apply.)***	1 No other adult observed or other adult did not talk	1 ☐
	2 English only	2 ☐
	3 Spanish or other language only	3 ☐
	4 Primarily English, some Spanish or other language	4 ☐
	5 Primarily Spanish or other language, some English	5 ☐
	6 English and Spanish (or other) about equally	6 ☐