Testing the Impact of Child Characteristics × Instruction Interactions on Third Graders' Reading Comprehension by Differentiating Literacy Instruction

Carol McDonald Connor
Florida Center for Reading Research, Florida State University, Tallahassee, USA

Fredrick J. Morrison, Barry Fishman
University of Michigan, Ann Arbor, USA

Sarah Giuliani, Melissa Luck, Phyllis S. Underwood, Aysegul Bayraktar, Elizabeth C. Crowe, Christopher Schatschneider
Florida Center for Reading Research, Florida State University, Tallahassee, USA

ABSTRACT

There is accumulating correlational evidence that the effect of specific types of reading instruction depends on children's initial language and literacy skills, called child characteristics × instruction (C×I) interactions. There is, however, no experimental evidence beyond first grade. This randomized control study examined whether C×I interactions might present an underlying and predictable mechanism for explaining individual differences in how students respond to third-grade classroom literacy instruction. To this end, we designed and tested an instructional intervention (Individualizing Student Instruction [ISI]). Teachers (n = 33) and their students (n = 448) were randomly assigned to the ISI intervention or a vocabulary intervention, which was not individualized. Teachers in both conditions received professional development. Videotaped classroom observations conducted in the fall, winter, and spring documented the instruction that each student in the classroom received. Teachers in the ISI group were more likely to provide differentiated literacy instruction that considered C×I interactions than were the teachers in the vocabulary group. Students in the ISI intervention made greater gains on a standardized assessment of reading comprehension than did students in the vocabulary intervention. Results indicate that C×I interactions likely contribute to students' varying response to literacy instruction with regard to their reading comprehension achievement and that the association between students' profile of language and literacy skills and recommended instruction is nonlinear and dependent on a number of factors. Hence, dynamic and complex theories about classroom instruction and environment impacts on student learning appear to be warranted and should inform more effective literacy instruction in third grade.

Students' ability to read and understand text is a key skill required for their academic and life success in our global and information-driven society. Yet, an alarming percentage of students, more than 70%, reach fourth grade unable to read and comprehend text at or above proficient levels, and this rate is higher for students who attend higher poverty schools (Lee, Grigg, & Donahue, 2007). Reading comprehension has been defined as the active extraction and construction of meaning from all kinds of text (Snow, 2002). One of the more important sources of influence on students' literacy development is the classroom instruction they receive (Morrison, Bachman, & Connor, 2005), and thus, finding ways to improve teachers' effectiveness with regard to reading comprehension instruction may prove a powerful tool in ensuring student achievement overall, especially for students living in poverty.

Whereas the field has been generally successful in identifying mechanisms for improving students' basic word reading skills, the anticipated growth in comprehension skills has not been realized (Gamse, Jacob, Horst, Boulay, & Unlu, 2008). One reason may be that, in general, teachers provide insufficient amounts of the types of instruction that are associated with stronger reading comprehension skill growth (Block, Parris, Reed, Whiteley, & Cleveland, 2009; Connor, Morrison, & Petrella, 2004; Improving America's Schools Act, 1994; National Institute of Child Health and Human Development [NICHD], 2000; Snow, 2002). Moreover, complicating teachers' task is accumulating evidence that the effect of particular types of instruction on reading gains may depend on students' reading and oral language skills (i.e., there are child characteristics by instruction (C×I) interactions, also called aptitude by treatment interactions (Connor et al., 2004; Connor, Jakobsons, & Granger, 2006; Connor, Piasta, et al., 2009; Cronbach & Snow, 1969; Juel & Minden-Cupp, 2000). Thus, specific instructional activities that are effective for students with typical reading and language skills may be ineffective for students with weaker or above-average skills and vice versa.

Although there is a general consensus in the educational community that differentiated reading instruction is a good thing (e.g., Tomlinson, 2001), there is surprisingly little empirical evidence or examination of the underlying mechanisms that might warrant such claims, particularly for reading comprehension. There is accumulating evidence that the impact of explicit instruction in the alphabetic principle, phonological awareness, and phonics depends on students' vocabulary and reading skills. This evidence is both correlational (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Juel & Minden-Cupp, 2000; Morrison et al., 2005) and experimental (Connor et al., in press; Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007). To date, however, there is virtually no experimental evidence of C×I interactions for the arguably more complex construct of third graders' reading comprehension, and only limited correlational evidence (Connor et al., 2004).

The purpose of this study was to explicitly consider whether C×I interactions represent an underlying mechanism that helps explain individual differences among third graders in their response to reading instruction as their reading skills move beyond basic decoding and increasingly toward reading for understanding. We did this by conducting a field experiment in which teachers and their students were randomly assigned to one of two interventions: one incorporating differentiated reading, called Individualizing Student Instruction (ISI), and the other incorporating an undifferentiated vocabulary intervention.

A Developmental Model of Reading Comprehension

As Perfetti, Landi, and Oakhill (2005) aptly noted, there are numerous reasons why many students have difficulty achieving proficient reading comprehension skills, which requires students to fluently decode and then understand what they are reading (Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007; Scarborough, 2001). Proficient reading comprehension is defined as the ability "to demonstrate an overall understanding of the text…to extend the ideas in the text by making inferences, drawing conclusions, and making connections to their own experiences" (National Assessment Governing Board, 2006, p. 24). Basic processes underlying reading comprehension are complex and call on the oral language system and a conscious understanding of this system (i.e., metalinguistic awareness) at all levels from semantic and morphosyntactic to pragmatic awareness (Morrison et al., 2005). Higher order metacognitive skills also appear to contribute to comprehension (Rapp et al., 2007; Willson & Rupley, 1997).

Thus, there is accumulating research on the underlying knowledge, skills, and strategies related to comprehension (NICHD, 2000; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001; Willson & Rupley, 1997). These include semantic knowledge and vocabulary (Biemiller & Boote, 2006), comprehension strategy use (NICHD, 2000; van den Broek, Risden, Fletcher, & Thurlow, 1996; Willson & Rupley, 1997), awareness of text structure (Williams, Stafford, Lauer, Hall, & Pollini, 2009), background knowledge (Rapp et al., 2007; Willson & Rupley, 1997), and self-regulation, including attention (McClelland et al., 2007).

Building on the work of Perfetti and colleagues (2005), Scarborough (1990), Catts and Kamhi (2004), and other researchers (e.g., Locke, 1993), as well as our own work (Connor, Piasta, et al., 2009), the current study relied on a developmental model of reading comprehension. The first assumption in this model is that the ability to read proficiently for understanding is built on students' developing social, cognitive, and linguistic systems. As these systems mature and increase in sophistication, so too do students' ability to co-opt these systems in the service of reading. In addition to decoding and letter/word reading skills, we consider comprehension processes, which may be largely automatic and unconscious higher order processes identified in the cognitive psychology literature (Perfetti, 2008; Rapp et al., 2007) or reflective or interrogative comprehension processes, which include conscious efforts to understand text and are largely identified in the education literature (NICHD, 2000; Pressley & Wharton-McDonald, 1997).

In this model, reading comprehension requires fluent decoding and word-level skills and fluent, automatic higher order processes, as well as the ability to use the automatic skills actively and consciously when the reading task demands it (i.e., reflective comprehension processes). The developmental model elucidates key skills that students bring to the task of learning that may moderate the impact of the reading instruction they receive on their comprehension gains (i.e., C×I interactions). These include students' basic word reading and decoding skills, their oral language, specifically vocabulary skills, and their comprehension skills.

Because reading comprehension is a complex construct, there is ongoing debate regarding the best way to assess reading comprehension (Keenan, Betjemann, & Olson, 2008; Sabatini & Albro, in press; Woodcock, McGrew, & Mather, 2001). In this study, we relied on well-regarded and psychometrically strong measures that are widely used in schools. The differentiated reading instruction intervention presented in this study used scores from the Woodcock-Johnson III (WJ–III) Passage Comprehension Test (Woodcock et al., 2001).

Although widely used, such assessments have been criticized by Keenan and colleagues (2008), who argued that the various comprehension assessments are not measuring the same skills. This conclusion is based on modest intercorrelations among the measures. Moreover, students' decoding and not their listening comprehension skills accounted for most of the variance on the WJ–III passage comprehension task. In our judgment, however, this cloze task requires students to utilize implicit understanding of the semantic and morphosyntactic systems to select the correct word that is missing from the sentence or passage. Thus, it assesses skills specifically identified in our developmental model. At the same time, we were cognizant of these concerns when we selected our outcome measure, the Gates–MacGinitie Reading Tests (GMRTs). The GMRT comprehension task, which requires students to read a fairly long passage and answer questions that are increasingly abstract (described more fully in the Methods section), arguably requires more inferencing and attention to text structure and has greater face validity than the WJ–III passage comprehension task, which is why we selected the GMRT comprehension task as our outcome for this study.

Characterizing Reading Instruction in Third Grade

The goal of reading instruction is to help students acquire the skills "that enable learning from, understanding, and enjoyment of written language" (Torgesen, 2002, p. 9). Increasingly, researchers and educators are finding that combinations of instructional activities and strategies are generally more effective than one method used to the exclusion of others (Block et al., 2009; NICHD, 2000). Indeed, when classrooms are observed, evidence reveals that effective teachers use a variety of strategies and types of lessons (Connor, Morrison, et al., 2009; Wharton-McDonald, Pressley, & Hampston, 1998). Moreover, with emerging evidence of C×I interactions, it is unlikely that we will find one single method of instruction that is optimal for all students. Thus, in this study, instruction is described across three dimensions focusing on (1) the content of the reading instruction, including phonological awareness, word decoding and encoding, text structure, vocabulary, and comprehension, (2) who is managing or focusing the students' attention on the learning activity at hand, the teacher or the student individually or with peers, and (3) grouping (i.e., whole class, small group, individual; Connor, Morrison, et al., 2009). These dimensions operate simultaneously to define any evidence-based literacy activity (see Table 1). We discuss each below.

Content of Instruction: Code- Versus Meaning-Focused Instruction

Content of instruction can be defined at different grain sizes from a fairly coarse curricular level (e.g., SRA/McGraw-Hill's Reading Mastery; Crowe, Connor, & Petscher, 2009) to a fairly fine level (e.g., teaching students to summarize). Following our developmental model of reading comprehension, in which proficient reading is a function of fluent decoding and strategic and flexible use of oral language (including semantic, morphosyntactic, and pragmatic skills) and background/academic knowledge to build coherent representations of the meanings of the text (Rapp et al., 2007), components of the content of literacy instruction can be defined. For this study, we use a coding system that examines instruction at a very fine grain, including types of morphemic awareness, types of listening and reading comprehension instruction (see Appendix A for definitions), to a larger grain size used in the ISI intervention, which we describe as either code- or meaning-focused instruction. The advantage of the larger grain size was that teachers were provided with more flexibility in selecting instructional activities based on

Table 1. Examples of Instructional Strategies Defined by Content and Management of Instruction

Dimension	Teacher/student-managed strategy	Student-managed strategy
Code-focused	The teacher works with a small group of students on an activity designed to help decode and spell multisyllabic words by using similar root words with different prefixes and suffixes. (morphological awareness)	Students work in small peer groups to practice spelling and decoding multisyllabic words. (word encoding)
Meaning-focused	The teacher, working with a small group of students, asks them to make inferences between two or more stories read in class in order to make connections and build background knowledge. (listening and reading comprehension)	Students work on a multiple-meaning vocabulary worksheet with the following words: *bark, story,* and *track* (print vocabulary). Other students engage in writing a summary of a story that they have recently read (writing).

their professional judgment and the scope and sequence of their core literacy curriculum.

Code-focused instruction is any instructional activity that builds students' grasp of the alphabetic principle, orthographic knowledge, and fluent decoding. This instruction includes phonics, phonological awareness, letter and word fluency, and spelling (see Table 1). In third grade, code-focused or word study instruction might include decoding multisyllabic words, morphological awareness, and other encoding strategies. Key is that code-focused instruction in third grade should likely focus on higher order and more complex decoding and encoding strategies than are observed in the earlier grades, depending on students' decoding skills.

Meaning-focused instruction is any instructional activity that is intended to improve students' ability to understand what they are reading and build coherent mental representations of the information in the text (Perfetti, 2008). Examples are provided in Table 1. Converging evidence reveals that from first through third grade, greater time in meaning-focused activities is associated with students' gains in reading comprehension (Connor, Piasta, et al., 2009; Guthrie et al., 2004; Williams et al., 2009). Meaning-focused activities include a wide range of activities, such as comprehension strategy instruction and practice, discussion, text reading, writing, and vocabulary, that may explicitly or implicitly affect reading comprehension gains.

Research has revealed that explicit instruction of comprehension strategies is associated with gains in reading comprehension and reading more generally (NICHD, 2000). Comprehension strategies include predicting, questioning, monitoring, highlighting, summarizing, using context clues, retelling, using prior knowledge, comparing and contrasting, and sequencing ideas (Block et al., 2009; NICHD, 2000; Pressley & Wharton-McDonald, 1997). In the direct and inferential mediation model (Cromley & Azevedo, 2007), for example, such comprehension strategies, in combination with students' decoding, oral language skills, and background knowledge, allow them to make appropriate inferences about the content of the text they are reading regarding information that is not explicitly stated in the text but can be inferred from information already conveyed in the text, their background knowledge, or other texts (Cain, Oakhill, & Lemmon, 2004). Of note, Cromley and Azevedo observed that students' ability to use strategies may not directly predict their comprehension but, instead, may support their ability to make inferences, which directly predicts their comprehension. Hence, in our coding system, we capture listening and reading comprehension instruction in great detail (see Appendix A).

The link between vocabulary and reading comprehension has been documented for over two decades, and correlational studies have shown a positive association between students' vocabulary knowledge and reading comprehension outcomes (Anderson & Freebody, 1981; Biemiller & Boote, 2006; Duke & Pearson, 2002; Storch & Whitehurst, 2002). Plus, vocabulary interventions have been associated with improved comprehension (Duke & Pearson, 2002; NICHD, 2000). The findings of the National Reading Panel, a meta-analysis of over 50 studies relating to best practices for the teaching of vocabulary instruction and its relation to reading comprehension, suggested that when instruction focused on building vocabulary, students' reading skills improved (NICHD, 2000). The National Reading Panel stated that "reading vocabulary is crucial to the comprehension processes of a skilled reader" (NICHD, 2000, p. 4-3).

Teacher/Student- Versus Student-Managed Instruction

An important dimension of instruction, but one that is frequently overlooked, is who is focusing students' attention on the learning activity at hand (Connor, Morrison, et al., 2009). Are the students working independently or with peers (i.e., student managed)? Or, is the teacher actively interacting with students and focusing their attention on the learning activity (i.e., teacher/student managed)? Examples of teacher/student-managed (TSM) and student-managed (SM) instruction are provided in Table 1. A teacher working with students to reach a consensus for the definition of a new vocabulary word is an example of a TSM activity. Students completing a vocabulary worksheet at their desks or in pairs are examples of SM activities.

Grouping: Whole-Class, Small-Group, and Individually Delivered Instruction

Another dimension of instruction captures the grouping context of instruction: whether it is delivered to all of the students in the classroom (i.e., whole class), to small groups of students, or individually. Many teachers use whole-class instruction, which is encouraged by several core literacy curricula (Block et al., 2009) and literacy approaches (Dahl & Freppon, 1995). However, research on effective schools (Wharton-McDonald et al., 1998) and differentiated or individualized instruction (Connor, Piasta, et al., 2009; Gersten et al., 2009) has indicated that the use of smaller, flexible learning groups based on students' current skills and learning needs may be more effective than whole-class instruction. Correlational evidence suggests that instruction provided in small groups may be up to four times as effective as instruction delivered to the entire class (Connor, Morrison, & Slominski, 2006). This is possibly because teachers may be more sensitive to students' response to what is being taught and can change instructional strategies and activities more flexibly to optimize learning.

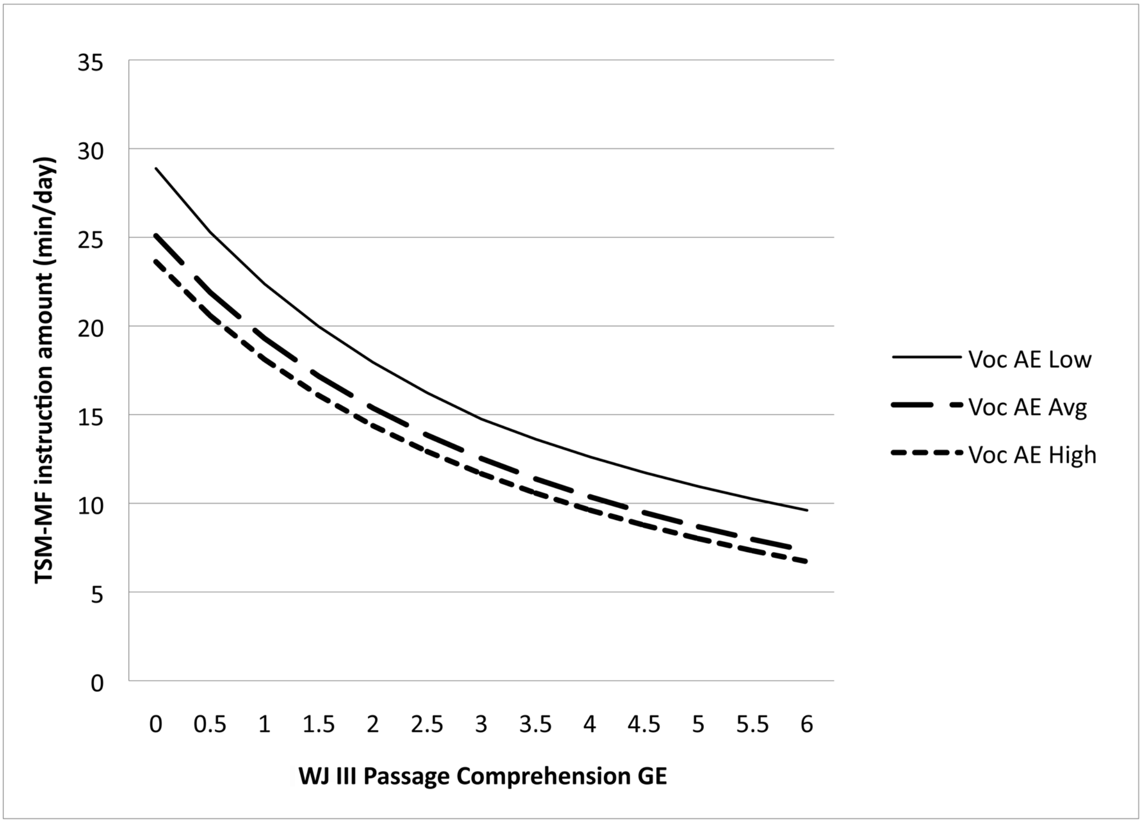
In this study, teachers in the ISI intervention group were specifically taught to use small, flexible learning groups based on students' reading skills while the other students worked in small peer groups or independently. During small-group time, teachers in the ISI intervention group were specifically taught to focus on providing instruction aligned with students' skills and abilities following recommendations based on C×I interaction research. The teachers in the vocabulary group were not specifically taught about using flexible learning groups, although of course, they were free to do so. Again, these dimensions operate simultaneously (Connor, Morrison, et al., 2009). Thus, the teacher working with a small group of students on decoding multisyllabic words would be defined as a TSM, small-group, code-focused activity (i.e., in the coding system, the content would be a type of morphological awareness). In the same way, a small group of students writing together in the publishing corner and discussing a story that they are writing would be a SM, small-group, meaning-focused activity.

C×I Interactions for Reading Comprehension

The correlational evidence for C×I interactions in third grade, although limited, has consistently shown that more time in explicit, TSM, meaning-focused (TSM-MF) types of instruction is associated with stronger student reading comprehension gains and that the effect is greater for students with weaker initial reading comprehension skills (Connor, Jakobsons, Crowe, & Meadows, 2009; Connor et al., 2004). At the same time, greater amounts of SM, code-focused (SM-CF) instruction appear to be associated with weaker gains in reading comprehension overall. Effective amounts of TSM, code-focused (TSM-CF) instruction depended on students' word decoding, vocabulary, and comprehension skills, and such instruction was generally only effective for students with word reading skills that fell below grade expectations (Connor et al., 2004). It is these C×I interactions that the present study was designed to test.

To test these C×I interactions, we created algorithms based on the hierarchical linear models (HLMs) used in the correlational studies. These HLM equations were reverse engineered, so to speak. The original equations could be used to predict a student's spring reading comprehension outcome based on fall scores, the amounts and types of reading instruction that the student received, and identified C×I interactions. To create the C×I algorithms, we set an outcome target, which we defined as on grade level by the end of the year or a school year's growth in reading comprehension, whichever was greater. Using grade equivalent (GE) as the metric, 3.9 would represent the minimum end of the year target in third grade. We then solved for amounts for each type of instruction (e.g., TSM-MF) using the student's assessed vocabulary, word reading, and passage comprehension GEs. The equations function somewhat like meteorologists' dynamical system forecasting models, which are used to predict, for example, the trajectory of hurricanes (National Hurricane Center, 2009). The key difference is that the models used in this study predict the amounts of each type of reading instruction that are required for students to reach their optimal trajectory of learning (Raudenbush, 2007) and have been called dynamical forecasting intervention models.

Figure 1. Recommended Minutes/Day of Teacher/Student-Managed, Meaning-Focused Instruction as a Function of Students' Woodcock-Johnson III Passage Comprehension Grade-Equivalent Score



Note. AE = age expectation. GE = grade equivalent. TSM-MF = teacher/student-managed, meaning-focused. Voc = vocabulary. WJ III = Woodcock-Johnson III. Students with vocabulary scores falling below AEs (top solid line, AE = 5 years) would be provided more time in TSM-MF instruction, for example, than would students with more typical vocabulary skills (middle dashed line, AE = 8.6 years, mean of the sample) or students with stronger vocabulary (bottom dotted line, AE = 11 years).

The equation used in the computer algorithm for TSM-MF instruction is provided in the following equation, and the recommended amounts charted as a function of students' reading comprehension GEs are provided in Figure 1.

$$\text{Recommended amount of TSM-MF instruction} = \frac{(8.3 \times \text{FRCGE}) - (\text{TO} - (0.75(\text{FVAE} + 8.0))) - 30}{-0.41(\text{FRCGE} - 1.74)} + 15$$

FRCGE = fall reading comprehension GE on the WJ III Passage Comprehension Test. FVAE = fall vocabulary age equivalent on the WJ–III Picture Vocabulary Test. TO = target outcome of 3.9 or (FRCGE + 0.9), whichever is greater.

As can be seen in Figure 1, the function is nonlinear and recommends exponentially more TSM-MF instruction as students' reading comprehension skills decrease. Relatively more time is recommended for students with weaker vocabulary scores. The recommended amounts are computed by the A2i (assessment to instruction) Web-based software, using these dynamical forecasting intervention models. A2i is described in the next section.

Interventions

Both interventions, ISI and vocabulary, were provided by the schools' general education classroom teachers as an integral part of the 90-minute block of time dedicated to literacy instruction during the 2008–2009 school year. All teachers used their school's core literacy curriculum, Open Court Reading, which encourages the use of small groups during workstation time and during the teaching of vocabulary. The interventions focused on improving teachers' practice for instruction that they were already expected to provide.

The ISI Intervention

The goal of the ISI intervention was to support teachers' efforts to differentiate reading instruction so that we could investigate the role of C×I interactions in understanding individual differences in students' literacy learning. The ISI intervention has three key components:

1. Assessment—All students receive vocabulary, word reading, and passage comprehension assessments three times per year, which are used in the A2i algorithms.
2. Assessment–instruction links that explicitly consider C×I interactions—The A2i software, which uses the dynamical forecasting intervention models (i.e., computer algorithms) to provide teachers with specific recommended amounts and types of literacy instruction for each student, computed using his or her vocabulary and reading scores, assessment and skill progress monitoring, and online training resources.
3. Professional development—Teachers' use of A2i and implementation of differentiated instruction in the classroom is supported through professional development provided by research-funded teacher mentors who are called research partners.

The professional development, coupled with the A2i software, is designed to provide teachers with explicit support and recommendations as they organize, plan, and differentiate literacy instruction. Teachers use their school's curriculum and other materials that they are currently using. Thus, ISI and A2i do not comprise a literacy curriculum per se. Rather, they provide a framework for differentiated instruction that relies on valid and ongoing assessment of three literacy skills—word reading, reading comprehension, and vocabulary knowledge—and empirical evidence regarding how instruction interacts with these skills to impact student outcomes. A2i might be described as an instructional decision support system (Landry, Anthony, Swank, & Monseque-Bailey, 2000), analogous to what is described in the medical field as a clinical decision support system (Garg et al., 2005; Kawamoto, Houlihan, Balas, & Lobach, 2005). In Response to Intervention parlance, the instruction provided would be considered differentiated Tier 1 or a hybrid Tier 1/Tier 2 intervention (Al Otaiba et al., in press; Gersten et al., 2009).

To access A2i (see isi.fcrr.org), teachers log on to the password-protected system and are taken to their homepage, where they can access text and video training materials (e.g., using assessment to guide instruction, using workstations and center activities effectively) and the planning components of A2i. The classroom view (see Figure 2) provides the recommended amounts of each type of literacy instruction (e.g., TSM-MF, TSM-CF) and recommended groupings based on students' reading comprehension skills. Teachers were encouraged to use the recommended groupings but could change them. Teachers were expected to provide the group mean recommended amounts in a flexible learning group format, attending to content and student skill level using their professional judgment. The literacy core curriculum, Open Court Reading, was indexed to the four types of instruction, as were the teacher-developed and Florida Center for Reading Research (www.fcrr.org) activities.

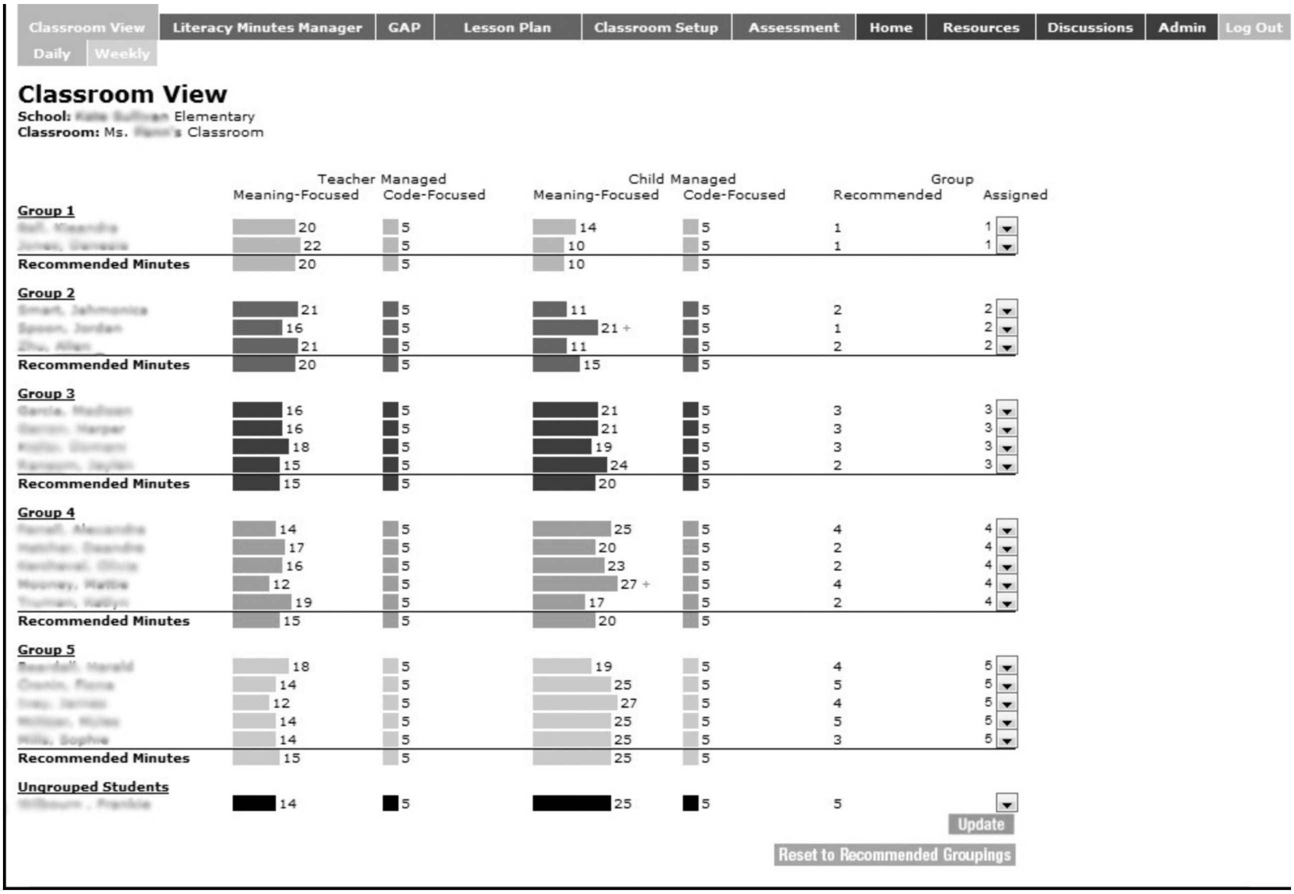
A computer screenshot showing a third-grade classroom with recommended amounts for each student is provided in Figure 2. Charts showing the algorithm-recommended amounts of instruction as a function of students' skills are provided in Figures 1 and 3. For students with generally weaker reading comprehension skills, more time daily in TSM-MF small-group instruction was recommended (see Figures 1 and 2), about 20 minutes per day for group 1, which included the students with the weakest reading and vocabulary scores. Fairly small amounts of SM, meaning-focused (SM-MF) instruction, about 10 minutes in the fall for group 1, with increasing amounts over the course of the school year were recommended.

For students with more typical reading comprehension skills (e.g., groups 3 and 4), about 15 minutes per day of TSM-MF, small-group instruction was recommended, with fairly substantial amounts of SM-MF instruction, about 20 minutes per day. For the strongest readers (group 5), about 10–15 minutes per day of TSM-MF and about 25 minutes per day of SM-MF instruction were recommended.

The A2i algorithms recommended very small amounts of TSM-CF instruction (5 minutes) unless students were reading well below grade expectations based on their word reading skills (see Figure 3). Indeed, in the classroom depicted in Figure 2, none of the students were reading more than half a grade below grade level, and five minutes per day were recommended for all. However, there were students in other classrooms for whom substantial amounts of TSM-CF instruction were recommended, and this amount increased exponentially as students' initial reading comprehension skills fell further below grade level. SM-CF instruction was held constant at five minutes per day, because there was no reliable correlational evidence regarding the contribution of this type of instruction to students' reading comprehension skills, with some indication of a negative association.

Teachers received intensive training using a coaching model (Bos, Mather, Narr, & Babur, 1999), which focused on how to use A2i software, implementing the recommended amounts and types of instruction in

Figure 2. A2i Classroom View Showing the Recommended Amounts (minutes/day) of Each Type of Instruction



	Teacher Managed Meaning-Focused	Teacher Managed Code-Focused	Child Managed Meaning-Focused	Child Managed Code-Focused	Group Recommended	Group Assigned
Group 1						
	20	5	14	5	1	1
	22	5	10	5	1	1
Recommended Minutes	20	5	10	5		
Group 2						
	21	5	11	5	2	2
	16	5	21 +	5	1	2
	21	5	11	5	2	2
Recommended Minutes	20	5	15	5		
Group 3						
	16	5	21	5	3	3
	16	5	21	5	3	3
	18	5	19	5	3	3
	15	5	24	5	2	3
Recommended Minutes	15	5	20	5		
Group 4						
	14	5	25	5	4	4
	17	5	20	5	2	4
	16	5	23	5	2	4
	12	5	27 +	5	4	4
	19	5	17	5	2	4
Recommended Minutes	15	5	20	5		
Group 5						
	18	5	19	5	4	5
	14	5	25	5	5	5
	12	5	27	5	4	5
	14	5	25	5	5	5
	14	5	25	5	3	5
Recommended Minutes	15	5	25	5		
Ungrouped Students						
	14	5	25	5	5	

Note. Student-managed code-focused instruction was set to a constant of five minutes per day. Teacher/student-managed, code-focused amounts depended on students' Woodcock-Johnson III letter/word identification grade-equivalent (GE) score (see Figure 3). None of the students in this classroom had letter/word GE scores that fell more than one GE below grade-level expectations.

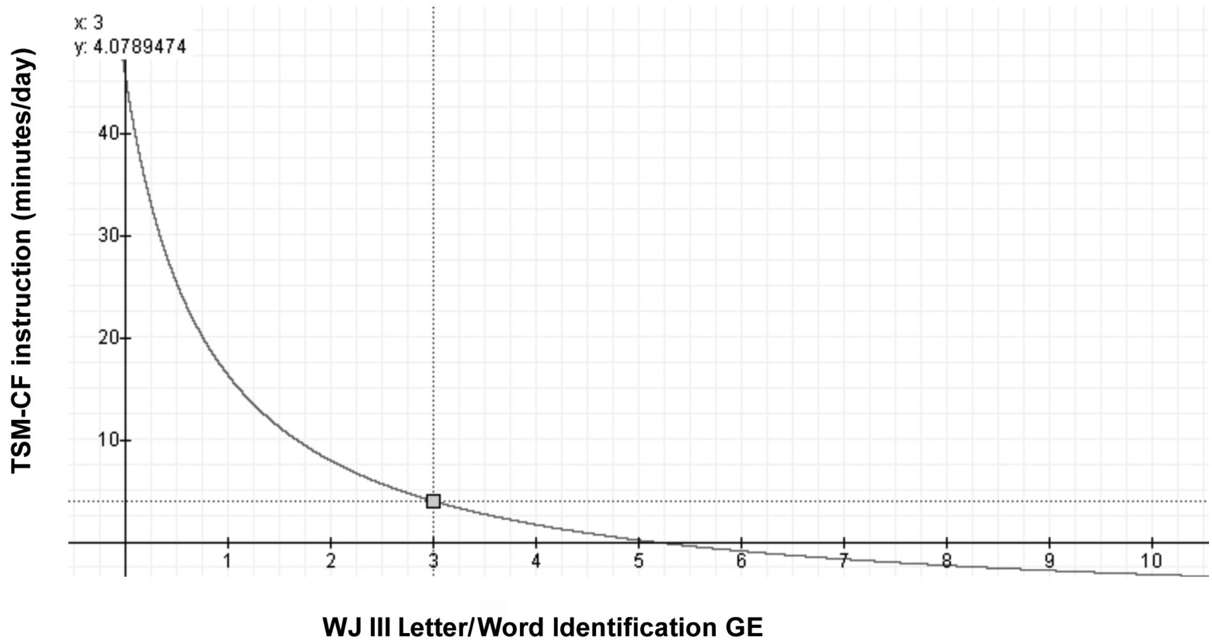
flexible learning groups. Other topics included planning and classroom management strategies and using assessment results to guide instruction. ISI group teachers participated in a half-day workshop in the fall, attended monthly one-hour meetings with other teachers in the ISI treatment group in communities of practice (Wenger, McDermott, & Snyder, 2002), and received biweekly classroom-based support during the literacy block (Bos et al., 1999).

The Alternative Vocabulary Treatment

For this study, the alternative vocabulary intervention focused implicitly on building students' comprehension by supporting teachers' efforts to provide effective vocabulary instruction using an adaptation of a teacher study group model (Bos et al., 1999; Gersten, Dimino, Jayanthi, Kim, & Santoro, 2007), in which teachers read the book *Bringing Words to Life: Robust Vocabulary Instruction* (Beck, McKeown, & Kucan, 2002). We selected this intervention because we wanted an intervention that might support students' vocabulary and comprehension growth but contrast with the ISI intervention, in that it was not differentiated by students' skill levels, nor was it intended to be. Prior to the monthly meetings, teachers read the assigned chapter in *Bringing Words to Life*. They then discussed the book during the meeting and, based on what they learned, designed vocabulary lessons collaboratively with a group of other teachers.

The grouping of teachers varied each month. Before the next meeting, each teacher implemented the lessons in his or her classroom. At the next meeting, the teachers discussed the implementation and shared student work samples. Then, the next chapter was discussed and new lessons developed. The procedure continued for each chapter throughout the school year. The research assistant leading the teacher study group was

Figure 3. Recommended Minutes/Day of Teacher/Student-Managed, Code-Focused Instruction as a Function of Students' Woodcock-Johnson III Letter/Word Identification Grade Equivalent



Note. GE = grade equivalent. TSM-CF = teacher/student-managed, code-focused. WJ III = Woodcock-Johnson III. A minimum of five minutes was set in the A2i software (see Figure 1).

a certified teacher working on her master's degree in reading and language arts.

In the book that the teachers read, discussed, and used as a guide for designing lessons, Beck et al. (2002) argued that for vocabulary instruction to be effective, instruction must be robust and explore information about target vocabulary words. The authors suggested that students' vocabulary will improve when teachers build students' background knowledge, provide them with multiple meanings of words across diverse contexts, and offer opportunities for them to read and listen to words. Such vocabulary instruction includes, but is not limited to, antonyms/synonyms, homonyms, classifying words, class discussion, and defining.

As described in *Bringing Words to Life*, words may have different levels of utility. The book describes three tiers of words: Tier 1, Tier 2, and Tier 3. Words that would be considered basic would be placed at Tier 1 (e.g., *baby, clock, happy, walk*). Words found in this tier rarely require instructional attention, according to Beck et al. (2002). Tier 1 also includes sight words and other words commonly found in a young student's environmental print and includes approximately 8,000 word families. Tier 2 words are considered high-frequency and high-utility words that are critical for understanding a specific text (Beck et al., 2002) and are used across multiple domains (e.g., *coincidence, absurd, industrious, fortunate*). Beck et al. suggested that a rich knowledge of Tier 2 words can have a positive impact on verbal functioning. Thus, instruction of Tier 2 words can be highly productive in supporting students' vocabulary and reading comprehension skills.

Beck et al. (2002) described Tier 3 words as occurring less frequently in written and spoken language and as genre specific, such as for science, math, and social sciences (e.g., *isotope, peninsula, lathe, refinery*). These words may be best learned when the need arises, such as introducing *peninsula* during a geography lesson, according to Beck et al. The professional development for the vocabulary intervention focused on supporting teachers' efforts to follow the Beck et al. approach to robust vocabulary instruction, as described in the book, as closely as possible.

Purpose of the Study

Again, the purpose of this study was to examine whether C×I interactions are causally implicated in students' varying reading comprehension outcomes in response to third-grade reading instruction. The research questions and hypotheses were the following:

- Research question 1—What is the effect of differentiating third-grade students' literacy instruction, using the ISI intervention, on their reading comprehension skill gains compared to the gains of students whose teachers were randomly assigned to the vocabulary intervention? If C×I instruction interactions are causally related to students' reading comprehension outcomes, we hypothesize that students whose teachers were in the ISI group would demonstrate stronger reading comprehension skill gains than would students whose teachers were in the vocabulary group. This is because, although both interventions seek to improve students' comprehension, the ISI intervention explicitly considers C×I interactions and the vocabulary intervention does not.
- Research question 2—What is the effect of the vocabulary intervention on students' vocabulary gains compared to gains for students in ISI classrooms? We anticipated that students whose teachers were in the vocabulary intervention group would demonstrate stronger vocabulary gains than would students in ISI classrooms.
- Research question 3—To further explore the role of C×I instruction interactions in students' learning, we asked, What was the nature and variability of the quality, amounts, and types of literacy instruction that third graders received during the dedicated block of time devoted to literacy? How precisely did teachers provide the A2i-recommended amounts and types of instruction? We anticipated that instruction would vary both within and between classrooms generally and that there would be systematic differences in amounts and types of instruction depending on the intervention condition. We predicted that students in ISI intervention classrooms would be more likely to receive the A2i-recommended amounts of small-group differentiated literacy instruction than would students in the vocabulary intervention classrooms. At the same time, we expected students in the vocabulary intervention classrooms to spend more time in oral language and print vocabulary instruction compared to students in ISI classrooms.

Methods

Participants

This study used a cluster-randomized treated control design in which third-grade teachers (n = 33) within schools (n = 7) were randomly assigned to implement the ISI intervention or implement an alternative treatment, the vocabulary intervention. All teachers were randomly assigned: 16 were assigned to the ISI group and 17 to the vocabulary group. No teachers withdrew; however, three teachers (n = 2 in the ISI group and 1 in the vocabulary group) went on leave and did not teach the last month of the study. These teachers and their students were included in all analyses. Similarly across groups, all teachers met state certification requirements. All 33 teachers reported that they had a bachelor's degree related to an educational field, and seven of the teachers had certifications or degrees beyond a bachelor's degree (n = 3 in the ISI group and 4 in the vocabulary group). Teachers' classroom teaching experience ranged from 0 to 30 years, with a mean of 10.9 years of experience (M = 11.2 years for the ISI teachers and 10.6 years for the vocabulary teachers). There were no significant differences in any of these teacher characteristics between the ISI intervention and vocabulary teachers.

The schools were located in a large school district in the southeastern United States and included suburban, urban, and rural communities. Schoolwide percentages of students qualifying for the federal free and reduced lunch program (FARL), which we used as a proxy for family socioeconomic status, ranged from high poverty (92%) to affluent (4%), with a mean of 47% of students qualifying for FARL studywide. The literacy blocks for these schools was about 90 minutes each day, and teachers used the school-adopted Open Court Reading curriculum (2008; Crowe et al., 2009).

Students were automatically assigned to the condition to which their teacher was assigned; thus, 219 students were in the ISI condition, and 229 were in the vocabulary condition (total n = 448). According to school records and parent reports, approximately 36% of the students were white, 51% were African American/black, 3% were Hispanic, 3% were Asian/Asian American, 3% were multiracial, and the remaining 4% belonged to other ethnic groups. There were no differences between the two intervention groups with regard to the percentage of students for race/ethnicity or qualifying for FARL. Of the 448 students, 100 had participated in ISI randomized control trials in first and second grade. Seven students were in control/alternative treatment classrooms for all three grades, 27 were in ISI classrooms for one of the grades, 42 were in ISI classrooms for two grades, and 24 were in ISI classrooms all three years. Comparison of fall reading and vocabulary total scores for this sample of 100 students with the sample as a whole indicates that their scores were not significantly different (p = .266). Fifty-eight students were in ISI classrooms and 42 in vocabulary classrooms.

Student Assessments

ISI Intervention Assessments

Students were assessed on a battery of language and literacy skills in fall, winter, and spring. GEs from the WJ–III Passage Comprehension, Letter/Word Identification, and Picture Vocabulary Tests (Woodcock et al., 2001) were used by the A2i software algorithms to compute recommended amounts and types of instruction. These scores were available to teachers in the ISI group throughout the school year. Paper reports of the scores were provided to teachers in the vocabulary condition.

Outcome Assessments

Reading comprehension and vocabulary were assessed in the fall and spring using the level 3 GMRTs, with alternative forms administered in the fall and spring. These scores were not used by the A2i software, although they were provided to teachers after administration and scoring were completed. These are multiple-choice, group-administered assessments. In the comprehension assessment, students read a variety of passages, including both narrative and expository text excerpted from books used widely in schools. Students then answer questions, with increasing levels of inference required, by selecting the best of four responses. For example, after reading a passage about emperor penguins, the students answer four questions, including, "Why does the mother penguin juggle the egg?"

In the reading vocabulary assessment, students select the meaning of a word provided in a short sentence (e.g., "a perfect grace") among four possible responses ("dance, grade, beauty of movement, lawn"). Internal consistency estimates (Cronbach's α) of 0.96 construct validity estimates, which show that the test is actually assessing the construct of reading comprehension, of about 0.80, and test–retest reliability ranging from 0.85 to 0.90 were reported for the 2006 standardization sample, which are acceptable (MacGinitie, MacGinitie, Maria, & Dreyer, 2006). Several types of scores are provided, including extended scale scores, GEs, percentile ranks, and normal curve equivalents. Extended scale scores, which have the advantage of providing equal intervals between points similar to a Rasch score (Winsteps version 3.30) to show gains in scores, were used in all analyses.

Instruction

For both conditions, instruction and the classroom environment were investigated in two ways: (1) using a rating scale from 1 (low) to 6 (high) with detailed rubrics designed to capture the general fidelity of teachers' implementation of ISI and vocabulary instruction on four scales that were specifically targeted to capture key aspects of the two interventions and which have been generally associated with more effective instruction (i.e., higher quality) in the extant literature (Brophy, 1979; Cameron, 2004; NICHD Early Child Care Research Network, 2004; Pianta, La Paro, Payne, Cox, & Bradley, 2002; Snow, Burns, & Griffin, 1998; Wharton-McDonald et al., 1998) and (2) capturing the amount and type of literacy instruction provided across three dimensions: management, context or grouping, and content.

Both systems relied on videotaped observations of instruction obtained during the literacy block in the fall, winter, and spring. Video was captured using two digital camcorders with wide-angle lenses. During the live observations, trained research assistants recorded detailed field notes regarding the activities and materials used, including careful descriptions of target students and activities of students who might be off camera (Bogdan & Biklen, 1998). Observations were scheduled at the teachers' convenience.

Fidelity

The fidelity of implementation was evaluated using four scales (see Appendix B):

1. Classroom implementation of individualized instruction—The extent to which teachers actually differentiated instruction in the classroom using small groups and centers, focusing on the content and level of the types of instruction (e.g., TSM-CF, SM-MF)
2. Classroom orienting, organization, and planning—The extent to which teachers planned center activities and small groups and used lesson plans (either their own or the A2i lesson plan) and in-classroom organizational strategies that supported effective and efficient use of instructional time (e.g., chart for students depicting centers and group membership)
3. Robust vocabulary instruction—The extent to which teachers provided robust vocabulary instruction
4. Warmth and responsiveness, control, and discipline—The extent to which teachers' classroom use of appropriate ways to redirect students' behavior, including warmth and responsiveness to students, supported effective and efficient instruction

Teachers received scores from 1 (low) to 6 (high) for each dimension. Each scale was considered separately for purposes of this study. The scales and rating rubrics are provided in Appendix B. Rubrics and scores were not shared with the participating teachers.

The scales were completed by research assistants who were certified teachers and, to the extent possible, blind to teachers' treatment group assignment. They observed video collected in late winter and early spring when, based on previous research, teachers are most likely to have mastered new ways of teaching (Hamre, Pianta, Downer, & Mashburn, 2007). Before beginning to rate instruction, coders worked together to achieve adequate levels of inter-rater reliability (Cohen's Kappa = 0.73). Approximately 10% of the coded videos were chosen at random and recoded. Inter-rater reliability remained at acceptable levels (Cohen's Kappa = 0.73) based on Landis and Koch (1977) criteria.

Amounts and Types of Instruction

Amounts and types of instruction were obtained using three classroom observation videos for each classroom: fall, winter, and spring. The instruction that each student received was coded across the three dimensions: management, grouping, and content (Connor, Morrison, et al., 2009). An excerpt of the coding manual is provided in Appendix A. Using Noldus Observer Video-Pro software (XT version 8.0), any activity (i.e., both instruction and noninstruction, e.g., transitions) that lasted at least 15 seconds was coded directly from video so that all of the instructional and noninstructional time that individuals spent during the literacy block was identified by content, management, and context. The output for one videotape from a classroom observed in spring 2009 is provided in Appendix C.

Literacy instruction was coded for a randomly selected subset of students (n = 364). Missing data analyses with fall and spring reading comprehension scores as the outcome revealed no significant differences between the full sample of 448 students and the selected sample of 364 students (Wilks's lambda = 0.996, F[2, 445] = 0.89, p = .411). For the selected students, 184 were in the ISI intervention classrooms and 180 in the vocabulary intervention classrooms.

For the purposes of this study, we focused on the amount and type of instruction provided individually or in small groups to students during the literacy block, as well as whole-class TSM instruction. Video coding was conducted by trained research assistants. All coders were required to attain acceptable inter-rater reliability (computed by the Noldus software) with a series of training videos of third-grade classrooms (Cohen's Kappa > 0.7). Inter-rater reliability during the coding process was obtained for about 10% of videos selected at random (mean Cohen's Kappa = 0.72), which is considered acceptable based on Landis and Koch (1977) criteria. The mean length of observation was 85 minutes in the fall (standard deviation [SD] = 30), 73 minutes in the winter (SD = 35), and 79 minutes in the spring (SD = 23).

The fine-grained coding system identifies over 200 instruction variables, and a sample of the coded output is provided in Appendix C. There were five code-focused types of instruction: phonological awareness, morpheme awareness, word decoding, word encoding, and fluency. There were six meaning-focused types of instruction coded: print and text concepts, oral language (including oral vocabulary), print vocabulary, listening and reading comprehension, text reading, and writing (spelling was coded as word encoding).

Within each content area, we considered specific activities. For example, in listening and reading comprehension, there were 19 different types of listening and reading comprehension activities (see Appendixes A and C). For this study, we considered only the duration of the principal content areas. Of note, instruction was coded at the level of the individual student (see Appendix C) rather than at the classroom level, unlike the quality ratings, which were global teacher/classroom-level ratings (see Appendix B). Thus, we could determine with some precision how much of each type of instruction each student received even if this varied within classrooms. Returning to the multiple dimensions of instruction, we considered whether the instruction was TSM or SM and whether it was provided to the whole class, in small groups, or individually.

To obtain a single value for each student, we aggregated the multiple observations first by summing results within season (for multiple videotapes) and then aggregated using the mean amount observed for each type of instruction for each student across the three observations.

To assess how precisely teachers provided the A2i-recommended amounts, difference scores were computed by subtracting the amount (in minutes) of each type of instruction observed (TSM-CF and -MF; SM-MF; small-group/individual instruction) from the A2i-recommended amount at the time of the observation. Thus scores that were closer to 0 indicated that the recommended amounts of each type of instruction were provided more precisely.

Analytic Strategies

For the first two research questions regarding the impact of the different interventions, because students were nested in classrooms that are nested in schools, we created a single model using HLMs (Raudenbush & Bryk, 2002), because failing to account for shared classroom and school variance may lead to misestimation of standard errors and, hence, effect sizes. We built the model systematically, starting with an unconditional model, which was used to compute intraclass correlations (ICCs). ICC is the classroom-level variance divided by the total model variance (student + classroom-level variance) and represents the variance explained between classrooms. Preliminary three-level models (students nested in classrooms that are nested in schools) revealed no significant between-school variance, so we used more parsimonious two-level models (students nested in classrooms) and included school-wide FARL. The model is provided here:

$$Y_{ij} = \beta_{0j} + \beta_{1j} \times \text{fall score}_{IJ} + r_{0ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} \times \text{treatment}_J + \gamma_{02} \times \text{school FARL}_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} \times \text{school FARL}_j$$

where Y_{ij} is the predicted spring score for child i in classroom j and is a function of the grand mean (γ_{00}), the students' fall score (γ_{10}), the effect of treatment (γ_{01})—where 1 = the ISI intervention, and 0 = the vocabulary intervention—and school FARL (γ_{02}) and the fall score × school FARL interaction (γ_{11}). r_{0ij} represents the student-level variance, and u_{0j} represents classroom-level variance.

For all analyses, continuous variables were grand mean centered.

Effect sizes were computed by dividing the coefficient of the treatment effect (γ_{01}) by the square root of the student-level variance, which is the standard deviation of the model outcome. Other analyses are described in the results.

To answer the third research question regarding the literacy instruction that the third graders received, we computed amounts of instruction (in seconds) that each student received for the fall, winter, and spring observations. For each season, students' data were aggregated by summing the observed amounts across multiple videotapes to obtain a single amount for each student in each of the major content area types for TSM and SM, small-group and whole-class instruction. To examine differences in instructional content (e.g., vocabulary, listening and reading comprehension) and the difference scores (i.e., precision) by condition, we used multilevel multivariate models (HMLM2; Raudenbush & Bryk, 2002; Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004).

Results

Across the conditions, students generally made grade-appropriate gains in reading comprehension from fall to spring when normal curve equivalents (NCEs), which should remain the same because they take into account students' age, were compared (fall reading comprehension NCE = 51.8; spring reading comprehension NCE = 50.9; $t[447] = 0.56$, $p = .574$, where NCEs have a standard mean of 50 and are similar to percentile ranks except that they have equal intervals and can be averaged). On the total score, students gained more than 2 NCE points, which was a significant increase, suggesting greater than expected growth overall on reading comprehension and vocabulary growth when combined (fall total NCE score = 52.7; spring total score NCE = 55.1; $t[447] = 4.85$, $p < .001$).

Extended scale scores' descriptive statistics are provided by condition in Table 2. Extended scale scores, which were used in the analyses, increased as expected from fall to spring. These scores are similar to raw scores except that they have equal intervals and have been scaled so that they can be used in statistical analyses and to model across grades. Comparison of fall scores for the ISI and vocabulary group students revealed that students in the ISI condition began the year with significantly lower reading comprehension and total scores on the GMRTs than did the students in the vocabulary condition ($t[446] = -2.98$, $p = .003$; $t[446] = -2.01$, $p = .045$, respectively). There were no significant differences between groups for fall reading vocabulary ($t[446] = -0.330$, $p = .742$).

Research Question 1

Three models were built using HLMs: one for the total GMRT score, one for comprehension, and another for reading vocabulary (see Table 3). Again, we

Table 2. Descriptive Statistics for the Gates–MacGinitie Comprehension, Reading Vocabulary, and Total Scores (extended scale scores) by Treatment Condition

Assessment group	Condition	Mean	Standard deviation
Fall total	Vocabulary	468.45	35.17
	ISI	461.27	40.34
	Total	464.94	37.92
Fall reading comprehension	Vocabulary	470.35	39.37
	ISI	458.77	43.01
	Total	464.69	41.55
Fall reading vocabulary	Vocabulary	468.14	37.93
	ISI	466.85	45.00
	Total	467.51	41.49
Spring total	Vocabulary	483.26	37.34
	ISI	480.47	37.58
	Total	481.90	37.44
Spring reading comprehension	Vocabulary	480.89	42.35
	ISI	476.25	43.27
	Total	478.62	42.82
Spring reading vocabulary	Vocabulary	487.75	37.91
	ISI	487.31	39.56
	Total	487.53	38.68

Note. ISI = Individualizing Student Instruction.

Table 3. Hierarchical Linear Model Results for Gates–MacGinitie Reading Tests (GMRTs)—Total Score, Reading Comprehension, and Reading Vocabulary Scores (extended scale scores)—Comparing Effects Between Treatment Conditions

Fixed effect	GMRTs' total coefficient	*p* value	Comprehension coefficient	*p* value	Vocabulary coefficient	*p* value
Fitted spring score	478.11	<.001	473.97	<.001	485.04	<.001
Student						
Fall score	0.81	<.001	0.76	<.001	0.74	<.001
Classroom						
ISI = 1	3.40	.049	4.95	.044	0.75	.724
School FARL	−0.14	<.001	−0.23	<.001	−0.14	<.001
Child × classroom interactions						
Fall score × school FARL	−0.001	.260	−0.001	.487	−0.001	.287
Random effects	**Variance**	***p* value**	**Variance**	***p* value**	**Variance**	***p* value**
Student ($r0_{1j}$)	316.51		640.48		396.97	
Classroom (u_{0j})	1.01	.180	20.88	.190	8.09	.095
Fall score (u_{1j})			0.01	.341		

Note. FARL = free and reduced lunch. ISI = Individualizing Student Instruction. ISI classroom = 1, and the vocabulary classroom = 0. Approximate degrees of freedom for fixed effects are 30 and 444 for GMRTs' total and vocabulary scores. Degrees of freedom for fixed effects are 30 and 31, respectively, for GMRTs' comprehension scores. Results control for the fall score and percentage of students qualifying for the school's FARL program. All continuous variables are grand mean centered.

hypothesized that if C×I interactions were causally implicated in students' varying reading comprehension outcomes in response to reading instruction, the ISI intervention would have a greater positive effect on students' reading comprehension skill growth than would the undifferentiated vocabulary intervention.

Supporting our hypothesis, HLM results revealed that students in the ISI intervention demonstrated significantly greater gains (i.e., residualized change) on the total GMRTs' score and on the reading comprehension assessment (see Table 3) compared with students who participated in the vocabulary intervention. The ISI effect size (d) for the total score was 0.19 and for reading comprehension was 0.20, which are relatively small effect sizes, using criteria suggested by Rosenthal and Rosnow (1984). The models explained 77% and 64% of the total variance, respectively. ICCs were 0.27 for the total score and 0.23 for comprehension. This indicates that 27% of the variability in students' total score and 23% of the variability in comprehension scores were explained by which classroom they attended. In all three models, as the schoolwide percentage of students qualifying for the school FARL program increased, students' outcome gains generally decreased. There was not a significant treatment × fall score interaction (comprehension × treatment coefficient = 0.053, $p = .304$; total GMRT × treatment coefficient = −0.036, $p = .547$).

Research Question 2

We anticipated that students whose teachers were in the vocabulary intervention group would demonstrate stronger vocabulary gains than would students in the ISI condition, but this was not supported by the results. HLM analyses revealed no significant differences in students' reading vocabulary gains (i.e., residualized change) whether their teachers were assigned to the vocabulary or the ISI intervention condition ($d = 0.04$; see Table 3). The model explained 73% of the total variance, and the ICC was 0.24. Nor was there a treatment × fall vocabulary score interaction (vocabulary × treatment coefficient = −0.001, $p = .301$).

Research Question 3

It is possible that the ISI intervention effect may have been the result of the professional development provided rather than the differentiated instruction informed by C×I interactions. To substantiate the claim that ISI represented instruction based on C×I interactions, we hypothesized that students in ISI intervention classrooms would be more likely to receive the A2i-recommended amounts of TSM, small-group, differentiated meaning-focused instruction and code-focused instruction than would students in the vocabulary intervention classrooms. We examined the overall fidelity of implemented instruction that students received to examine qualitative differences in fidelity as an explanation for the ISI treatment effect. We then examined amounts and types of instruction that each group received and compared the precision with which teachers provided the A2i-recommended amounts, but first we examined the amounts and types of instruction that third graders received overall in these classrooms.

Overall Description of Third-Grade Literacy Instruction

Amounts (in seconds) of each type of instruction for TSM, small-group and individual instruction are provided in Table 4 and for TSM, whole-class instruction, aggregated to the classroom level, in Table 5. For these third graders, very little TSM-CF, small-group instruction was observed compared with TSM-MF, small-group instruction: less than 3 minutes per day (SD = 5) in TSM-CF, small-group activities, compared with more than 20 minutes per day (SD = 17) spent in meaning-focused activities. Amounts of small-group instruction ranged widely among classrooms from 0 to 50 minutes of TSM-CF instruction during the literacy block and 0 to almost 75 minutes of TSM-MF instruction.

Overall, about three times as much time was spent in TSM-CF, whole-class instruction (13 minutes/day, SD = 11, range = 0–44 minutes) as in small-group and individual instruction. Even more time was spent in TSM-MF, whole-class instruction (53 minutes/day, SD = 27, range = 14–155 minutes).

Generally, TSM-CF activities tended to focus on the more complex skills of morphological awareness, word decoding, and word encoding. Very little time was spent on phonological awareness and grapheme–phoneme correspondences. Most of the time in TSM-MF, small-group instruction was spent in text reading (about 7 minutes) or listening and reading comprehension activities (6 minutes). Only about 2 minutes per day were spent in writing activities. TSM-MF, whole-class instruction was generally spent in listening and reading comprehension, text reading, and print vocabulary instruction.

Adding together small-group, individual, and whole-class instruction indicated that students spent about 35 minutes per day (SD = 22), on average, in TSM and SM literacy instruction, but this ranged from as little as 2 minutes for one student to more than 105 minutes for another. This does not include time spent in transition and other noninstructional activities or time that students were not in the classroom.

Table 4. Means and Standard Deviations in Seconds/Day of Teacher/Student-Managed, Small-Group and Individual Instruction by Content Area for 347 Target Students

Content area	Condition	Mean	Standard deviation
Phonological awareness	Vocabulary	17.69	58.23
	ISI	1.74	9.91
	Total	9.42	41.73
Morphological awareness	Vocabulary	32.51	100.76
	ISI	129.19	294.87
	Total	82.66	228.46
Word identification and decoding	Vocabulary	11.66	33.92
	ISI	23.96	72.59
	Total	18.04	57.58
Word identification and encoding	Vocabulary	4.49	40.32
	ISI	71.00	154.97
	Total	38.99	119.63
Grapheme–phoneme correspondence	Vocabulary	5.48	32.51
	ISI	8.74	64.36
	Total	7.17	51.50
Fluency	Vocabulary	14.80	58.31
	ISI	33.27	66.88
	Total	24.38	63.49
Print and text concepts	Vocabulary	49.98	208.45
	ISI	50.79	158.93
	Total	50.40	184.16
Oral language	Vocabulary	1.53	6.95
	ISI	20.43	45.43
	Total	11.34	34.35
Print vocabulary	Vocabulary	136.41	277.71
	ISI	236.05	401.12
	Total	188.10	350.32
Listening and reading comprehension	Vocabulary	292.74	281.04
	ISI	428.83	425.44
	Total	363.33	369.00
Text reading	Vocabulary	335.70	590.06
	ISI	481.59	472.19
	Total	411.38	536.40
Writing	Vocabulary	218.07	353.52
	ISI	79.14	189.22
	Total	146.00	288.64

Note. ISI = Individualizing Student Instruction.

Comparing Instruction for the Two Conditions

Fidelity

Again, the fidelity of implementation was rated for four aspects of the classroom environment: individualization, organization and planning, robust vocabulary instruction, and teacher warmth and responsiveness at the level of the classroom. We hypothesized that teachers in the ISI group would receive higher ratings on the individualization and organization/planning scales, whereas teachers in the vocabulary group would receive higher ratings for robust vocabulary. We assumed that there would be no differences in teacher warmth and responsiveness. For all four scales, ratings for the ISI

Table 5. Means and Standard Deviations in Seconds/Day of Teacher/Student-Managed, Whole-Class Instruction by Content Area

Content area	Condition	Mean	Standard deviation
Phonological awareness	Vocabulary	94.63	126.75
	ISI	42.50	98.55
	Total	68.56	114.78
Morphological awareness	Vocabulary	255.13	303.94
	ISI	278.38	432.35
	Total	266.75	367.82
Word identification and decoding	Vocabulary	192.69	215.13
	ISI	178.88	171.48
	Total	185.78	191.50
Word identification and encoding	Vocabulary	157.75	259.22
	ISI	244.31	336.36
	Total	201.03	298.65
Grapheme–phoneme correspondence	Vocabulary	34.81	51.96
	ISI	30.88	41.30
	Total	32.84	46.21
Fluency	Vocabulary	102.31	171.95
	ISI	62.94	85.93
	Total	82.63	135.20
Print and text concepts	Vocabulary	349.75	467.25
	ISI	188.75	148.97
	Total	269.25	350.81
Oral language	Vocabulary	150.25	184.34
	ISI	100.50	110.53
	Total	125.38	151.63
Print vocabulary	Vocabulary	906.56	925.52
	ISI	625.69	479.04
	Total	766.13	738.83
Listening and reading comprehension	Vocabulary	1065.00	599.10
	ISI	779.75	737.10
	Total	922.38	676.44
Text reading	Vocabulary	919.44	489.26
	ISI	751.50	470.33
	Total	835.47	479.73
Writing	Vocabulary	119.94	166.97
	ISI	229.94	471.47
	Total	174.94	352.38

Note. ISI = Individualizing Student Instruction.

and vocabulary intervention classrooms did not significantly differ (see Table 6).

We used multivariate ANOVA (MANOVA, using PASW version 17.0.3), because fidelity was judged at the level of the classroom, so there was no nesting. Results revealed that there were no overall significant differences in fidelity of observed instruction between teachers in the two intervention groups (Wilks's lambda = 0.880, $F[4, 28] = 1.137$, $p = .359$). Post hoc analyses revealed that ISI and vocabulary intervention teachers did not differ significantly on any of the four scales with effect sizes all negligible (partial eta squared ≤ 0.07). Differences might have existed, but we did not have the power to detect them. The smallest effect size that could be detected given the teacher sample size and parameters of the model was 0.30 (GPower version 3.1). In general, trends were in the direction anticipated, with teachers in the ISI intervention group demonstrating slightly higher mean scores on the individualization and planning and organization scales and the vocabulary intervention group teachers demonstrating slightly higher mean scores on the vocabulary scale. Generally, teachers received the highest ratings on the planning and organization scale and the lowest ratings for individualized instruction.

Amount of Each Type of Instruction

Two-level HLMs (Raudenbush & Bryk, 2002) examining TSM, small-group code- and meaning-focused instruction (minutes/day) revealed substantial within- and between-classroom variability. HLMs were used because small-group instruction was observed at the level of the individual student, and hence students were nested in classrooms. For TSM-CF, small-group instruction, within-classroom variance (i.e., student level, r) was 1.32, and the between-classroom variance (u_0) was 2.07 ($\chi^2[31] = 519.45$, $p < .001$). The ICC was 0.61, which indicates that more than half of the variability in students' amount of instruction received was explained by which classroom they attended. For TSM-MF, small-group instruction, the within-classroom variance (r) was 15.66, and the between-classroom variance (u_0) was 70.13 ($\chi^2[31] = 1688.03$, $p < .001$). The ICC was 0.81. Both ICCs represent very high levels of between-classroom variance (Hedges & Hedberg, 2007).

Multivariate multilevel analyses (HMLM2; Raudenbush et al., 2004) were used to take into consideration the nested structure of the observation data, individual students nested in classrooms, and the significant between-classroom variance in the amounts of each instruction type (see Table 7). The multiple variables included each type of TSM, small-group and individual instruction, with code- and meaning-focused models run separately to preserve parsimony for these highly complex models. For both models, the unrestricted model provided the best fit (see Raudenbush & Bryk, 2002). Results revealed that students in the ISI intervention spent significantly more time overall in TSM, small-group and individual meaning- and code-focused instruction compared with students in the vocabulary intervention group.

Comparing ISI and vocabulary classroom amounts of TSM, whole-class instruction (see Table 5 for means) using MANOVA revealed that students in ISI and vocabulary classrooms generally spent the same amount of time in whole-class instruction (Wilks's lambda = 0.705, $F[12, 19] = 0.663$, $p = .765$). MANOVA is appropriate because whole-class instruction is aggregated to the classroom level, and the data do not have a nested structure. Examining between-subjects effects for TSM, whole-class instruction by content area revealed no mean differences by content area with partial eta squared ranging from 0.001 to 0.053. Thus, teachers in both groups appeared overall to be providing comparable amounts of whole-class literacy instruction to students.

Precision

The ISI intervention A2i software specifically provided recommended amounts (in minutes/day) of small-group or individual instruction, reflecting predicted C×I interactions. These were used to compute difference scores. Again, difference scores were calculated by subtracting a student's A2i-recommended minutes from the student's observed minutes of small-group and individual instruction for each instruction type (e.g., TSM-MF). Hence, difference scores closer to 0 indicated that the student received more precisely the recommended amounts of each type of instruction (see Table 8). If the student received less than the recommended amount, the student's difference score was negative. If the student received more than the recommended amount, the difference score was positive. Because the recommended amounts varied by month and were recalculated after the winter assessments, we computed the difference score for each season separately and then computed the mean difference score for each student. Means, standard deviations, and ranges are provided in Table 8.

HLM analyses, controlling for the fall reading comprehension score (sample grand mean centered) and schoolwide percentage of students qualifying for FARL, revealed that TSM-CF difference scores were closer to 0 by about one minute, on average, and hence more precise for the ISI intervention students compared with the vocabulary intervention students (see Table 9).

Teachers in the ISI intervention were also more precise in providing the recommended amounts of TSM-MF, small-group instruction than were teachers in the vocabulary intervention. That is, students in the ISI group received instruction that was generally closer to the A2i-recommended amounts (i.e., more precise) by about four minutes per day than was the instruction received by the vocabulary group. However, the precision of TSM-MF, small-group instruction increased when students' fall reading comprehension scores were greater (see Figure 4). There were no significant differences in the SM-MF difference scores across conditions.

Table 6. Means and Standard Deviations for Teacher Fidelity Quality Scales by ISI and Vocabulary Intervention Groups

Scale	Condition	Mean	Standard deviation
Individualized instruction	Vocabulary	2.32	1.58
	ISI	2.94	1.91
	Total	2.62	1.75
Organization and planning	Vocabulary	4.32	1.17
	ISI	4.91	1.00
	Total	4.61	1.12
Use of robust vocabulary strategies	Vocabulary	3.71	1.17
	ISI	3.56	1.14
	Total	3.64	1.14
Warmth and responsiveness to students, control, and discipline	Vocabulary	4.26	1.46
	ISI	4.38	1.06
	Total	4.32	1.26

Note. 1 = low, and 6 = high. ISI = Individualizing Student Instruction. There were no significant differences by condition overall or by scale.

Table 7. Results of Multilevel Multivariate Models Comparing Teacher/Student-Managed (TSM), Meaning- and Code-Focused, Small-Group and Individual Instruction for Students in ISI and Vocabulary Intervention Classrooms

Variable	TSM, meaning-focused instruction		TSM, code-focused instruction	
Fixed effects	Coefficient	*p* value	Coefficient	*p* value
Fitted mean (intercept)	–2.33	.657	14.44	<.001
ISI = 1	14.95	.045	30.21	<.001
Random effect τ	Variance		Variance	
	340.75		1309.36	

Note. ISI = Individualizing Student Instruction. Degrees of freedom for intercept and ISI = 30.

Using HLMs, we then examined the association between precision of TSM-MF, small-group instruction and reading comprehension outcomes. We included all students because there was likely some drift of the ISI intervention, and notably, some vocabulary teachers used small-group instruction. Results, controlling for schoolwide FARL and fall reading comprehension scores, revealed that the more precisely teachers provided the A2i-recommended amounts (i.e., difference score closer to 0 minutes), the greater were students' reading comprehension gains (coefficient = 0.734, standard error = 0.02, $p = .001$). The effect size (d) for the range of difference scores for just students in the ISI classrooms (21.95 − 0 minutes) was 0.64.

Table 8. Difference Scores (minutes/day) by Condition

Type of instruction	Condition	Mean	Standard deviation	Minimum	Maximum
TSM code-focused difference score	Vocabulary	−5.08	1.37	−11.71	0.28
	ISI	−4.28	2.23	−11.30	8.15
TSM meaning-focused difference score	Vocabulary	−9.18	8.81	−22.39	29.36
	ISI	−6.14	10.56	−21.94	18.38
SM meaning-focused difference score	Vocabulary	−11.17	7.13	−22.97	14.33
	ISI	−10.34	6.78	−25.08	8.25

Note. Difference scores closer to 0 indicate greater precision meeting the A2i recommended amounts. ISI = Individualizing Student Instruction. SM = student-managed. TSM = teacher/student-managed.

Table 9. Hierarchical Linear Model Results Comparing Difference Scores by Condition (ISI = 1; vocabulary = 0), Controlling for Fall Reading Comprehension and School Free and Reduced Lunch

Variable	TSM, code-focused difference score coefficient	TSM, code-focused difference score standard error	TSM, meaning-focused difference score coefficient	TSM, meaning-focused difference score standard error
Intercept or fitted mean	−5.08***	0.33	−9.85***	2.20
Student level				
Fall RC	0.009**	0.003	0.03***	0.009
Classroom level				
ISI	0.97*	0.47	4.12	3.10
School FARL	0.005	0.007	−0.02	0.05
Child × classroom				
Fall RC × ISI	0.001	0.004	0.03*	0.01
Fall RC × FARL			−0.0003	0.0002
Random effects	Variance	Chi-square	Variance	Chi-square
Classroom	1.58	262.93	75.32***	1,479.08
Student	2.05		17.58	

Note. FARL = free and reduced lunch. ISI = Individualizing Student Instruction. RC = reading comprehension. TSM = teacher/student-managed. All continuous variables are grand mean centered.
*$p < .05$. **$p < .01$. ***$p \le .001$.

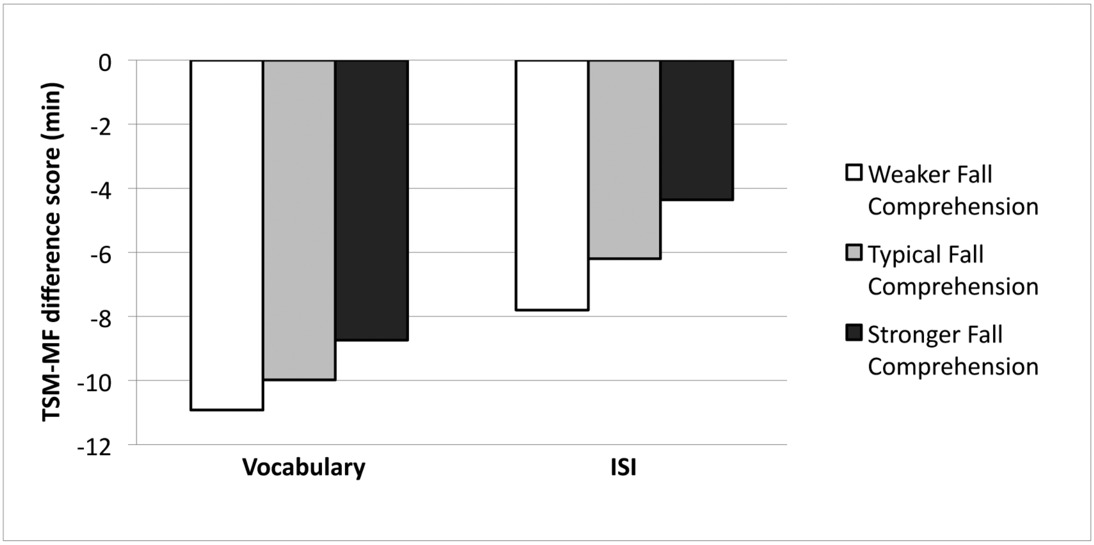
Discussion

The results of this study revealed that the ISI intervention designed to explicitly consider C×I interactions was generally more effective in improving students' reading comprehension than was instruction of similar quality that did not take into account C×I interactions. Supporting our hypothesis, we found that, on average, students in the ISI condition demonstrated greater gains in reading comprehension overall than did students in the vocabulary condition. Moreover, teachers in the ISI condition, as compared with the vocabulary condition, were more likely to provide the amounts of each type of small-group and individual instruction recommended by A2i, as evidenced by significantly smaller differences between the observed and recommended amounts of instruction (i.e., difference scores that were generally closer to 0).

Together with other experimental and quasi-experimental evidence (Al Otaiba et al., in press; Connor et al., 2007; Connor et al., in press), this study extends to third grade the accumulating evidence that C×I interactions are likely causally implicated in students' reading achievement and their response to the classroom instruction that they receive. Hence, current calls to differentiate classroom instruction (Gersten et al., 2009) and Response to Intervention initiatives are likely appropriate in light of these results.

Our hypothesis regarding the vocabulary intervention, in which teachers received professional

Figure 4. Difference Scores for Students in the Vocabulary and ISI Classrooms as a Function of Their Fall Reading Comprehension Scores



Note. Fall comprehension = Fall Gates–MacGinitie Reading Tests extended scale scores (ESSs). ISI = Individualizing Student Instruction. TSM-MF = teacher/student-managed, meaning-focused. Reading comprehension is modeled at the 25th (white = 435 ESS), 50th (gray = 462 ESS), and 75th (black = 489 ESS) percentiles of the sample. Difference scores closer to 0 indicate greater precision in the A2i recommended amounts of TSM, meaning-focused, small-group instruction that students received (observed amount – A2i recommended amount).

development on providing robust vocabulary instruction, was not supported. There were no significant between-group differences in students' reading vocabulary gains. Of note, when components of instruction were considered in post hoc analyses, teachers in the vocabulary group were not more likely to provide either small-group or whole-class instruction in oral language or print vocabulary instruction than were ISI teachers. Moreover, when we examined the overall quality of instruction, there were no significant differences in quality when comparing the ISI and vocabulary teachers' vocabulary instruction.

Indeed, across all four scales—individualizing instruction, planning and organizing, vocabulary, and teacher warmth and responsiveness to students—we found that, in general, teachers in both conditions provided fairly high–quality instruction with fairly strong ratings (average 4.6 out of possible 6) for planning and instruction and teacher warmth and responsiveness, control, and discipline (4.3 out of 6). There were no significant differences between the ISI and vocabulary intervention teachers; however, with global ratings for only 33 teachers, we lacked the power to reject the null hypothesis with certainty. The trends for each scale were in the expected direction based on the intervention to which teachers were assigned. Teachers in the ISI intervention tended to achieve higher quality ratings for individualizing and planning, whereas teachers in the vocabulary intervention tended to achieve higher ratings for vocabulary instruction. Ratings were close to equivalent for teacher warmth and responsiveness, control, and discipline, which was integral to both interventions. This may be one reason for the relatively small treatment effect ($d = 0.20$ for reading comprehension). The ISI treatment effect likely represents an effect over and above generally high-quality instruction.

The evidence for C×I interactions in combination with findings in kindergarten and first grade (Al Otaiba et al., in press; Connor, Piasta, et al., 2009) indicate that we cannot assume that a one-size-fits-all whole-class instructional approach promoted in many core literacy curricula is going to be generally effective for many third graders, especially for students who begin third grade with very strong or very weak skills. As we define *high-quality instruction*, we have to ask for which student with which profile of skills and consider that these profiles are changing over time. What is effective and high-quality instruction for one student may be ineffective and, hence, poor quality for a student with a different profile of skills. These considerations have implications with regard to assessment and aligning instruction with students' assessed skills.

The role of assessment in designing effective instruction is likely an active ingredient to supporting effective differentiated instruction (Cohen, Raudenbush, & Ball, 2003), as is making the links between assessment and instruction more explicit. Additionally, assessing higher order language and comprehension skills appears to be a crucial part of planning effective early elementary literacy instruction. Students' oral language and literacy skills consistently interact with instruction types from kindergarten through third grade (Al Otaiba et al., in press; Foorman et al., 2006; Juel & Minden-Cupp, 2000). At the same time, there was generally a nonlinear association between students' initial skill profile and recommended amounts of instruction (see Figure 1). This means that how to differentiate instruction is arguably more complex and less intuitive than the current practice of using benchmarks, such as Response to Intervention, might warrant. With valid and reliable initial and ongoing assessment of key skills, in this case word reading, reading comprehension, and vocabulary, and better understanding of students' skill profiles, teachers and specialists should be better able to design and implement effective literacy instruction by taking into account C×I interactions.

Although we provided assessment results to teachers regardless of their assigned condition, the ISI teachers had access to A2i software, which as its name implies, explicitly links assessment results to recommendations for amounts and types of literacy instruction. Teachers could view graphs that showed each student's progress as well as progress for the entire class. Plus, the A2i-recommended instruction amounts and groups were directly tied to current assessment results for each student. Access to such salient links have been implicated in other education and medical research for improving student and patient outcomes (Connor, Piasta, et al., 2009; Kawamoto et al., 2005; Landry et al., 2009). Important, teachers in the ISI condition were trained to differentiate the level and presentation of the content during TSM, small-group time. Teachers received training in planning, classroom organization, using assessment, including informal in-the-moment assessment to guide instruction, and the use of evidence-based literacy activities. We conjectured that the small-group context may be ideally suited to supporting teachers' ability to make fine-grained instructional decisions and align the content and delivery of the instruction that they provide with students' instructional needs.

Because we found the anticipated group differences in the precision with which teachers provided the A2i-recommended amounts, it is likely that all of the components of ISI—assessment, use of the A2i software, providing the recommended amounts in small groups with differentiated level and content, and professional development—contributed to the significant positive impact of the ISI intervention. A limitation of any randomized controlled trial is identifying specific active ingredients of multicomponent interventions. To the extent possible, we attempted to keep the interventions similar. Both interventions were implemented during the regularly scheduled and daily dedicated block of time to literacy instruction, and all teachers received professional development, assessment information, and technical support.

Conversely, only the ISI teachers used the A2i software and were directly supported in their efforts to differentiate literacy instruction in line with C×I interactions. Our classroom observations provide evidence that the ISI teachers were more precise than the vocabulary teachers in providing recommended amounts. Additionally, the magnitude of students' gains increased with increased TSM-MF instruction precision. Thus, it is likely that providing TSM, meaning- and code-focused, small-group instruction that was based on C×I interactions was an active ingredient, which likely contributed to the treatment effect. This most directly supports our claim that C×I interactions are implicated in students' varying responses to even high-quality instruction.

Real-world classroom observations provide data that allow researchers to investigate the nature of instruction, activities, and materials (Fraenkel & Wallen, 2009). For example, we observed that participating third-grade teachers were more likely to provide more complex code-focused instruction than has been generally observed in first grade (Foorman et al., 2006). When code-focused instruction was provided, it was more likely to cover morphological awareness, encoding, decoding, and fluency than the more basic skills of phonological awareness and grapheme–phoneme correspondence. Additionally, much more instructional time was spent in meaning-focused activities than in code-focused activities.

What is central about the observation system used in this study is that instruction was coded for individual students. Demario (a pseudonym for student B; see Appendix C) might be reading a book with the teacher while Samantha (a pseudonym for student L) is finishing a comprehension assignment at her desk. This study, as well as previous research, demonstrates that students who share the same classroom do not receive the same instructional and learning opportunities (Tuyay, Jennings, & Dixon, 1991) whether by design or not. This observation system also considers multiple dimensions of instruction so that content (i.e., both instruction and noninstruction), grouping, and management of focused attention (i.e., both TSM and SM instruction) are captured simultaneously. Arguably, literacy is a multidimensional construct (Rapp et al., 2007), hence examining instruction across multiple dimensions is likely to be more informative than more global classroom indicators alone. At the same time, by adding global indicators of instructional quality (Pianta, Belsky, Houts, & Morrison, 2007), researchers can begin to explicate the aspects of the classroom learning environment (e.g., duration, quality) that actively contribute to students' language and literacy learning.

Our classroom observation findings illustrate the importance of TSM-MF instructional strategies during third grade with decreasing amounts of code-focused instruction provided, except for students who continue to demonstrate difficulties with basic word reading skills. Particularly, TSM-MF instruction, in which the teacher provides explicit, differentiated instruction to a small group of students, appears to be effective for many students, including those who begin third grade with reading comprehension skills falling below or above grade expectations. However, our results also indicate that students with weaker fall reading comprehension scores were less likely to receive the A2i-recommended amounts. There are a few possible explanations. First, students with weak reading comprehension skills were also likely to have weaker decoding skills (in this sample, $r = 0.45$, $p < .001$); hence, teachers may have preferred to spend small-group time focused on code-focused skills rather than meaning-focused skills. It is possible that the importance of providing small-group, TSM-MF instruction was not well understood by the teachers. Finally, teachers may be less comfortable providing effective TSM-MF instruction in small groups than they are with TSM-CF instruction. More research is needed.

Neither intervention had a significant effect on students' reading vocabulary gains. We had hypothesized that students whose teachers participated in the vocabulary intervention would show greater gains on our vocabulary measures than would students in the ISI group, but this was not the case. In other studies (Gersten et al., 2007), the teacher study group protocol using the Beck et al. (2002) book has had a positive effect on students' vocabulary. It may be that had we targeted a specific lexicon, we would have observed growth on the specific words taught. Many vocabulary interventions do not make a significant difference on standardized tests of vocabulary (NICHD, 2000), and trajectories of growth on these measures are fairly stable among students, although absolute scores vary.

With 33 teachers and 448 students, lack of power is a limitation of this study. Using optimal design (Raudenbush & Liu, 2000), post hoc analyses indicated that the smallest effect size (d) that we could reliably detect (power = 0.8) was 0.40. By adding covariates, particularly students' initial status and schoolwide FARL, we were able to increase power to detect an effect size of 0.20. Nevertheless, null findings should be interpreted cautiously, particularly teacher-level analyses in which power was less. Although randomized control trials are among the most robust designs for establishing causal relations (Shavelson & Towne, 2002), any school-based study is messy, and all findings should be interpreted cautiously until replicated. Attrition is among the most serious concerns. Although we had no attrition, some teachers went on leave for the last month of the study (their data were included in the analyses). Additionally, there was variability in the fidelity with which teachers implemented the two interventions, as evidenced by the fairly large standard deviations displayed in Tables 4–6 and 8. It is quite likely that effect sizes might have been larger with stronger fidelity. Finally, we randomly assigned teachers within schools, which improved power but may have contributed to drift of the intervention. It is possible that teachers shared strategies, so aspects of the ISI intervention may have influenced instruction in the vocabulary classrooms and visa versa. This would have the tendency to decrease the size of the treatment effects.

The significant effect of treatment on a well-regarded standardized test of reading comprehension was obtained by regular classroom teachers who began ISI training and use of A2i software in the fall of the observed school year. They were able to change and sustain this change in their literacy instruction practices following the ISI intervention protocol so that their students made greater gains overall in reading comprehension compared with students whose teachers received the vocabulary intervention's professional development but were not taught how to use assessment to differentiate student instruction. The ISI teachers were able to change and sustain this change even though many teachers report that using assessment information to inform differentiated reading instruction can be challenging (Roehrig, Duggar, Moats, Glover, & Mincey, 2008). Professional development methods used were those generally found in the extant literature to be effective in helping teachers improve their practice and have been used by many schools (Bos et al., 1999; Chard, 2004; Fishman, Marx, Best, & Tal, 2003; Ingvarson, Meiers, & Beavis, 2005). One difference is that teachers had access to technology designed to support their practice (Kawamoto et al., 2005). These findings offer some possible suggestions for supporting teachers' efforts to improve their practices.

Results also highlight the nonlinear and dynamic nature of classroom literacy instruction in the way that it affects students' learning (Yoshikawa & Hsueh, 2001). The association between students' initial profile of language and literacy skills and recommended amounts is nonlinear (see Figures 1 and 3) and, for some types of instruction, changes each month. Moreover, as students' skills change, so too do the recommended amounts of instruction. Dynamic forecasting intervention models provide a concrete way to help teachers interpret complex assessment data and design instruction for students with varying profiles of word reading, reading comprehension, and vocabulary skills. As we strive to improve students' ability to read for understanding by considering the role of C×I interactions, using data to inform instruction, and conceptualizing instruction across multiple dimensions more dynamically and proactively, we can improve students' literacy achievement, including their reading comprehension skills.

Notes

This study was funded by grants R305H04013 and R305B070074, "Child by Instruction Interactions: Effects of Individualizing Instruction," from the Institute of Education Sciences of the U.S. Department of Education and by grant R01HD48539 from the Eunice Kennedy Shriver NICHD. The opinions expressed are ours and do not represent views of the funding agencies. We thank Tricia Rizza, Stephanie Day, Jennifer Dombek, Novel Tani, and the ISI Project team members for their hard work in providing professional development, collecting data, and coding videotapes. We thank Dr. Stephanie Al Otaiba and Dr. Young Suk Kim for their advice on a recent version of this manuscript. Additionally, we thank the students, parents, teachers, and school administrators without whom this research would not have been possible.

References

Al Otaiba, S., Connor, C.M., Folsom, J.S., Greulich, L., Meadows, J., & Li, Z. (in press). Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal*.

Anderson, R.C., & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association.

Beck, I.L., McKeown, M.G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York: Guilford.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44–62. doi:10.1037/0022-0663.98.1.44

Block, C.C., Parris, S.R., Reed, K.L., Whiteley, C.S., & Cleveland, M.D. (2009). Instructional approaches that significantly increase reading comprehension. *Journal of Educational Psychology, 101*(2), 262–281. doi:10.1037/a0014319

Bogdan, R.C., & Biklen, S.K. (1998). *Qualitative research in education: An introduction to theory and methods* (3rd ed.). Boston: Allyn & Bacon.

Bos, C.S., Mather, N., Narr, R.F., & Babur, N. (1999). Interactive, collaborative professional development in early literacy instruction: Supporting the balancing act. *Learning Disabilities Research & Practice, 14*(4), 227–238. doi:10.1207/sldrp1404_4

Brophy, J.E. (1979). Teacher behavior and its effects. *Journal of Educational Psychology, 71*(6), 733–750. doi:10.1037/0022-0663.71.6.733

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*(4), 671–681. doi:10.1037/0022-0663.96.4.671

Cameron, C.E. (2004, April). *Variation in teacher organization*. Paper presented at the 18th biennial Conference on Human Development, Washington, DC.

Catts, H.W., & Kamhi, A.G. (Eds.). (2004). *Language and reading disabilities* (2nd ed.). Boston: Allyn & Bacon.

Chard, D.J. (2004). Toward a science of professional development in early reading instruction. *Exceptionality, 12*(3), 175–191. doi:10.1207/s15327035ex1203_5

Cohen, D.K., Raudenbush, S.W., & Ball, D.L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119–142. doi:10.3102/01623737025002119

Connor, C.M., Jakobsons, L.J., Crowe, E.C., & Meadows, J.G. (2009). Instruction, student engagement, and reading skill growth in Reading First classrooms. *The Elementary School Journal, 109*(3), 221–250.

Connor, C.M., Jakobsons, L.J., & Granger, J. (2006). *Individual differences in the reading fluency of children and their instructional implications: Instruction and fluency*. Paper presented at the 14th annual meeting of the Pacific Coast Research Conference, Coronado, CA.

Connor, C.M., Morrison, F.J., Fishman, B.J., Ponitz, C.C., Glasney, S., Underwood, P.S., et al. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher, 38*(2), 85–99. doi:10.3102/0013189X09332373

Connor, C.M., Morrison, F.J., Fishman, B.J., Schatschneider, C., & Underwood, P. (2007). Algorithm-guided individualized reading instruction. *Science, 315*(5811), 464–465. doi:10.1126/science.1134513

Connor, C.M., Morrison, F.J., & Petrella, J.N. (2004). Effective reading comprehension instruction: Examining child X instruction interactions. *Journal of Educational Psychology, 96*(4), 682–698. doi:10.1037/0022-0663.96.4.682

Connor, C.M., Morrison, F.J., Schatschneider, C., Toste, J., Lundblom, E.G., Crowe, E., et al. (in press). Effective classroom instruction: Implications of child characteristic by instruction interactions on first graders' word reading achievement. *Journal of Research on Educational Effectiveness*.

Connor, C.M., Morrison, F.J., & Slominski, L. (2006). Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology, 98*(4), 665–689. doi:10.1037/0022-0663.98.4.665

Connor, C.M., Piasta, S.B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., et al. (2009). Individualizing student instruction precisely: Effects of child X instruction interactions on first graders' literacy development. *Child Development, 80*(1), 77–100. doi:10.1111/j.1467-8624.2008.01247.x

Cromley, J.G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*(2), 311–325. doi:10.1037/0022-0663.99.2.311

Cronbach, L.J., & Snow, R.E. (1969). *Individual differences in learning ability as a function of instructional variables*. Stanford, CA: Stanford University. (ERIC Document Reproduction Service No. ED029001)

Crowe, E.C., Connor, C.M., & Petscher, Y. (2009). Examining the core: Relations among reading curricula, poverty, and first through third grade reading achievement. *Journal of School Psychology, 47*(3), 187–214. doi:10.1016/j.jsp.2009.02.002

Dahl, K.L., & Freppon, P.A. (1995). A comparison of innercity children's interpretations of reading and writing instruction in the early grades in skills-based and whole language classrooms. *Reading Research Quarterly, 30*(1), 50–74.

Duke, N.K., & Pearson, P.D. (2002). Effective practices for developing reading comprehension. In A.E. Farstrup & S.J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed., pp. 205–242). Newark, DE: International Reading Association.

Fishman, B.J., Marx, R.W., Best, S., & Tal, R.T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*(6), 643–658. doi:10.1016/S0742-051X(03)00059-3

Foorman, B.R., Francis, D.J., Fletcher, J.M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to

read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*(1), 37–55. doi:10.1037/0022-0663.90.1.37
Foorman, B.R., Schatschneider, C., Eakin, M.N., Fletcher, J.M., Moats, L.C., & Francis, D.J. (2006). The impact of instructional practices in grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology, 31*(1), 1–29. doi:10.1016/j.cedpsych.2004.11.003
Fraenkel, J.R., & Wallen, N.E. (2009). *How to design and evaluate research in education* (7th ed.). New York: McGraw-Hill.
Gamse, B.C., Jacob, R.T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First Impact Study: Final report* (NCEE 2009-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
Garg, A.X., Adhikari, N.K.J., McDonald, H., Rosas-Arellano, M.P., Devereaux, P.J., Beyene, J., et al. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *The Journal of the American Medical Association, 293*(10), 1223–1238. doi:10.1001/jama.293.10.1223
Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2009). *Assisting students struggling with reading: Response to Intervention (RtI) and multi-tier intervention in the primary grades* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
Gersten, R., Dimino, J., Jayanthi, M., Kim, J., & Santoro, L. (2007, July). *Impact of teacher study groups on observed teaching practice and student vocabulary and comprehension for first grade teachers: Results of large scale randomized controlled trials.* Paper presented at the 14th annual meeting of the Society for the Scientific Study of Reading, Prague, Czech Republic.
Guthrie, J.T., Wigfield, A., Barbosa, P., Perencevich, K.C., Taboada, A., Davis, M.H., et al. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology, 96*(3), 403–423. doi:10.1037/0022-0663.96.3.403
Hamre, B.K., Pianta, R.C., Downer, J., & Mashburn, A.J. (2007, April). *Growth models of classroom quality over the course of the year in preschool programs.* Paper presented at the biennial meeting of the Society for Research in Child Development, Boston.
Hedges, L.V., & Hedberg, E.C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60–87. doi:10.3102/0162373707299706
Improving America's Schools Act of 1994, Pub. L. No. 103-382, § 1201 (1994).
Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes and efficacy. *Education Policy Analysis Archives, 13*(10).
Juel, C., & Minden-Cupp, C. (2000). Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly, 35*(4), 458–492. doi:10.1598/RRQ.35.4.2
Kawamoto, K., Houlihan, C.A., Balas, E.A., & Lobach, D.F. (2005). Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *British Medical Journal, 330*(7494), 765–768. doi:10.1136/bmj.38398.500764.8F
Keenan, J.M., Betjemann, R.S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*(3), 281–300. doi:10.1080/10888430802132279
Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
Landry, S.H., Anthony, J.L., Swank, P.R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology, 101*(2), 448–465. doi:10.1037/a0013842
Lee, J., Grigg, W., & Donahue, P. (2007). *The nation's report card: Reading 2007: National Assessment of Educational Progress at grades 4 and 8* (NCES 2007-496). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved February 27, 2011, from nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007496
Locke, J.L. (1993). *The child's path to spoken language.* Cambridge, MA: Harvard University Press.
MacGinitie, W.H., MacGinitie, R.K., Maria, K., & Dreyer, L.G. (2006). *Gates-MacGinitie reading tests* (4th ed.) [Kit]. Itasca, IL: Riverside.
McClelland, M.M., Cameron, C.E., Connor, C.M., Farris, C.L., Jewkes, A.M., & Morrison, F.J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*(4), 947–959. doi:10.1037/0012-1649.43.4.947
Morrison, F.J., Bachman, H.J., & Connor, C.M. (2005). *Improving literacy in America: Guidelines from research.* New Haven, CT: Yale University Press.
National Assessment Governing Board. (2006). *Reading framework for the 2007 National Assessment of Educational Progress.* Washington, DC: U.S. Department of Education.
National Hurricane Center. (2009). *Technical summary of the National Hurricane Center track and intensity models* (Rev. ed.). Miami, FL: National Hurricane Center, National Centers for Environmental Prediction, NOAA/National Weather Service. Retrieved February 27, 2011, from www.nhc.noaa.gov/model summary.shtml
National Institute of Child Health and Human Development. (2000). *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* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.
NICHD Early Child Care Research Network. (2004). Multiple pathways to early academic achievement. *Harvard Educational Review, 74*(1), 1–29.
Open Court Reading [Curriculum]. (2008). Washington, DC: IES What Works Clearinghouse. Retrieved July 19, 2009, from ies.ed.gov/ncee/wwc/reports/beginning_reading/open_court/
Perfetti, C.A. (2008, April). *Reading comprehension: A conceptual framework from word meaning to text meaning.* Paper presented at the Assessing Reading in the 21st Century Conference, Philadelphia.
Perfetti, C.A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M.J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Malden, MA: Blackwell.
Pianta, R.C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science, 315*(5820), 1795–1796.
Pianta, R.C., La Paro, K.M., Payne, C., Cox, M.J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal, 102*(3), 225–238.
Pressley, M., & Wharton-McDonald, R. (1997). Skilled comprehension and its development through instruction. *School Psychology Review, 26*(3), 448–466.
Rapp, D.N., van den Broek, P., McMaster, K.L., Kendeou, P., & Espin, C.A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11*(4), 289–312.

Raudenbush, S.W. (2007, April). *Distinguished Contributions to Education Research Award (2006) lecture: How shall we study the cause and effects of classroom teaching?* Paper presented at the annual meeting of the American Educational Research Association, Chicago.
Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
Raudenbush, S., Bryk, A., Cheong, Y., Congdon, R., & du Toit, M. (2004). *HLM 6: Hierarchical linear and nonlinear modeling.* Lincolnwood, IL: Scientific Software International.
Raudenbush, S.W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods, 5*(2), 199–213.
Rayner, K., Foorman, B.R., Perfetti, C.A., Pesetsky, D., & Seidenberg, M.S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*(2), 31–74. doi:10.1111/1529-1006.00004
Roehrig, A.D., Duggar, S.W., Moats, L., Glover, M., & Mincey, B. (2008). When teachers work to use progress monitoring data to inform literacy instruction: Identifying potential supports and challenges. *Remedial and Special Education, 29*(6), 364–382. doi:10.1177/0741932507314021
Rosenthal, R., & Rosnow, R.L. (1984). *Essentials of behavioral research: Methods and data analysis.* New York: McGraw-Hill.
Sabatini, J., & Albro, E. (Eds.). (in press). *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences.* Lanham, MD: Rowman & Littlefield.
Scarborough, H.S. (1990). Very early language deficits in dyslexic children. *Child Development, 61*(6), 1728–1743.
Scarborough, H.S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S.B. Neuman & D.K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York: Guilford.
Shavelson, R.J., & Towne, L. (Eds.). (2002). *Scientific research in education.* Washington, DC: National Academy Press.
Snow, C.E. (2002). *Reading for understanding: Toward a research and development program in reading comprehension.* Santa Monica, CA: RAND.
Snow, C.E., Burns, M.S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.
Storch, S.A., & Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934–947. doi:10.1037/0012-1649.38.6.934
Tomlinson, C.A. (2001). *How to differentiate instruction in mixed-ability classrooms* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
Torgesen, J.K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40*(1), 7–26. doi:10.1016/S0022-4405(01)00092-9
Tuyay, S., Jennings, L., & Dixon, C. (1991). Classroom discourse and opportunities to learn: An ethnographic study of knowledge construction in a bilingual third-grade classroom. *Discourse Processes, 19*(1), 75–110. doi:10.1080/01638539109544906
van den Broek, P., Risden, K., Fletcher, C.R., & Thurlow, R. (1996). A "landscape" view of reading: Fluctuating patterns of activation and the construction of a stable memory representation. In B.K. Britton & A.C. Graesser (Eds.), *Models of understanding text* (pp. 165–187). Mahwah, NJ: Erlbaum.
Wenger, E., McDermott, R., & Snyder, W.M. (2002). *Cultivating communities of practice: A guide to managing knowledge.* Boston: Harvard Business School Press.
Wharton-McDonald, R., Pressley, M., & Hampston, J.M. (1998). Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *The Elementary School Journal, 99*(2), 101–128.
Williams, J.P., Stafford, K.B., Lauer, K.D., Hall, K.M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology, 101*(1), 1–20.
Willson, V.L., & Rupley, W.H. (1997). A structural equation model for reading comprehension based on background, phonemic, and strategy knowledge. *Scientific Studies of Reading, 1*(1), 45–63. doi:10.1207/s1532799xssr0101_3
Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement.* Itasca, IL: Riverside.
Yoshikawa, H., & Hsueh, J. (2001). Child development and public policy: Toward a dynamic systems perspective. *Child Development, 72*(6), 1887–1903. doi:10.1111/1467-8624.00384

Submitted July 17, 2010
Final revision received January 31, 2011
Accepted February 11, 2011

Carol McDonald Connor is an associate professor of psychology at the Florida Center for Reading Research at Florida State University, Tallahassee, USA; e-mail cconnor@fsu.edu.

Fredrick J. Morrison is a professor of psychology at the University of Michigan, Ann Arbor, USA; e-mail fjmorris@umich.edu.

Barry Fishman is an associate professor of learning technologies at the University of Michigan, USA; e-mail fishman@umich.edu.

Sarah Giuliani is a doctoral student at the Florida Center for Reading Research at Florida State University, USA; e-mail seg08d@fsu.edu.

Melissa Luck is a doctoral student at the Florida Center for Reading Research at Florida State University, USA; e-mail mjr7024@fsu.edu.

Phyllis S. Underwood is a postdoctoral fellow and project director at the Florida Center for Reading Research at Florida State University, USA; e-mail punderwood@fcrr.org.

Aysegul Bayraktar is a research assistant at the Florida Center for Reading Research at Florida State University, USA; e-mail aysegulbayraktar@yahoo.com.

Elizabeth C. Crowe is a postdoctorate fellow and project director at the Florida Center for Reading Research at Florida State University, USA; e-mail eccrowe@gmail.com.

Christopher Schatschneider is a professor in psychology at the Florida Center for Reading Research at Florida State University, USA; e-mail schatschneider@fsu.edu.

Appendix A

Excerpts From the ISI Classroom Observation System Coding Manual

7.1.8 Oral Language (Behavior)

Oral Language should only be coded for activities that do not involve print. The intent of the activity is to increase students' oral vocabularies (i.e., their ability to access a word's meaning upon hearing it pronounced) and/or listening and speaking abilities. Activities intended to expand students' knowledge of word meanings but where the print form of the word is given or displayed are better coded under Print Vocabulary. Activities intended to increase students' comprehension skill are better coded under Comprehension. Activities in which the meanings of multimorphemic words are deduced by analyzing the meanings of the individual morphemes should be coded under Morpheme Awareness > Structural Analysis.

7.1.8.2 Vocabulary/Teacher/Student Defines (Modifier)

Oral Language > Vocabulary/Teacher/Student Defines should be coded when the activity involves the teacher/student giving the definition of a word. The word is not seen in its print form when its definition is being given.

7.1.8.3 Vocabulary/Class Discussion (Modifier)

Oral Language > Vocabulary/Class Discussion should be coded when the activity involves a class discussion to arrive at a word's meaning; the class discusses the word, or a number of students give various definitions until a consensus is reached. For example, the teacher asks the class or small group for the definition of *vacant*, and multiple students give/attempt the definition. The word is not seen in its print form when its definition is being given. An example would be when one student's definition does not include the connotation of the word (e.g., defining *donate* as "give"), and the teacher calls on other students to expand on the definition (e.g., "giving to charity or as a gift"). Oral Language > Vocabulary/Class Discussion is often seen when students are defining words with slightly different shades of meaning.

7.1.8.4 Vocabulary/Use (Modifier)

Oral Language > Vocabulary/Use should be coded when the activity involves understanding the pragmatic and semantic use of a word (e.g., using words in sentences; explaining how, when, and where a particular word would be used). The word is not seen in its print form when its definition is being given.

7.1.8.5 Classifying Words (Modifier)

Oral Language > Classifying Words should be coded for activities in which the students make semantic maps or are listing like words. For example, the teacher asks the class to list words that are a type of fruit. The print form of the word is not seen until after it is listed.

7.1.8.6 Antonym (Modifier)

Oral Language > Antonym should be coded when the activity involves generating or matching words with opposite meanings or discussing the concept of an antonym. The print forms of the words are not given during the activity.

7.1.8.7 Synonym (Modifier)

Oral Language > Synonym should be coded when the activity involves generating or matching words with similar meanings or discussing the concept of a synonym. The print forms of the words are not given during the activity. The intent of the activity is not to define a particular word (e.g., "what does *donate* mean?" with a student answer of "give"); these activities should be coded under the relevant Oral Language > Vocabulary code.

7.1.8.8 Homonym (Modifier)

Oral Language > Homonym should be coded for activities that involve words that sound similar or when discussing the concept of a homonym; the words may or may not differ in spelling (i.e., homophones like *bear/bare*; also, *bear* meaning an animal that lives in the woods or to yield or carry). The difference in the words' meanings is made explicit. The print forms of the words are not given during the activity.

7.1.8.9 Pragmatics (Modifier)

Oral Language > Pragmatics should be coded for activities in which students are expected to consider the role of audience, purpose of speaking, and so forth in an example of oral language (e.g., a speech). If the discussion of pragmatics is limited to a single word, this should be coded under Oral Language > Vocabulary/Use.

7.1.8.10 Sharing (Modifier)

Oral Language > Sharing should be coded when students and/or the teacher are talking about personal business to the group. This category includes activities such as show-and-tell or sharing during a morning meeting, as well as when the class is sitting in a circle and just chatting with the teacher.

Sharing should involve both teacher and student input, and it should be clear that both expect the other to speak in a reciprocal way. A student who bursts out with news from home during calendar time is interrupting and should be coded Noninstruction > Disruption, unless the teacher takes that opportunity to ask whether other students have things to share and continues the conversation with the class (Pathways code). Discussion that involves brainstorming for a writing activity should be coded as Writing > Prewriting/Discussion.

7.1.8.11 Sentence Expansion (Modifier)

Oral Language > Sentence Expansion should be coded for activities that involve increasing sentence complexity by adding adverbs, adjectives, nouns, and so forth for the purpose of teaching vocabulary. If the intent of the activity is just to learn how to increase sentence complexity by adding adjectives, adverbs, verbs, and so forth, then this should be coded as Print Text Concepts > Sentence Expansion. Or, if the focus of the activity is on teaching students to write their own complex sentences, then this should be coded as Writing > Sentence Expansion.

7.1.8.12 Context Cues (Modifier)

Oral Language > Context Cues should be coded for activities when the teacher explicitly explains, models, or prompts students to use context cues to aid in determining the meaning of an unidentified word (e.g., using pictures and surrounding words/text). If text is present, then this activity should be coded as Print Vocabulary > Context Cues. This should not be confused with Word ID Decoding/Encoding > Context Cues when the purpose is to identify the pronunciation of an unidentified word, or with Comprehension > Context Cues when the purpose is to use pictures or surrounding text to understand a new event or information presented in the text.

7.1.10 Listening and Reading Comprehension (Behavior)

Comprehension should be coded for activities intended to increase students' comprehension of written or oral text. This includes instruction and practice in using comprehension strategies, and demonstration of comprehension abilities. Comprehension activities generally follow or are incorporated into reading of or listening to connected text (e.g., silent sustained reading followed by a comprehension worksheet, comprehension strategy instruction using a particular example of connected text, or an interactive teacher read-aloud during which the teacher models various comprehension strategies).

7.1.10.2 Previewing (Modifier)

Listening and Reading Comprehension > Previewing should be coded for activities that involve thinking about what might occur in a story based on the illustrations (including taking a picture walk through a book), cover, title, and so forth. Previewing activities always *precede* reading and involve predictions about the general content of a text, which helps distinguish it from Comprehension > Predicting. Previewing often leads into activating prior knowledge related to the story (Comprehension > Prior Knowledge).

7.1.10.3 Schema and Concept Building (Modifier)

Listening and Reading Comprehension > Schema Building should be coded for activities that involve the *teacher* clarifying a concept and building background knowledge. For example, the teacher tells the students about the Middle Ages while reading a fairy tale. Discussions about specific words should be coded as Print Vocabulary > Class Discussion.

7.1.10.4 Predicting (Modifier)

Predicting should be coded for activities that involve predicting future events or information not yet presented based on information already conveyed by the text (e.g., making predictions from foreshadowing). Predicting occurs while reading a story and involves specific details or events, as opposed to Comprehension > Previewing, which involves a general prediction of what the text will be about.

7.1.10.5 Inferencing—Between Texts (Modifier)

Inferencing—Between Texts should be coded for activities that involve making inferences between two or more stories in a text based on information that is not explicitly stated in the text but is inferred from information already conveyed in the text. An example of this would be if the teacher is reading a story about a boy who loses his dog and then the teacher tells the students, "Remember the story we read last week about the boy who lost his favorite hat? How did that boy feel? Do you think the boy in this story feels any different?"

7.1.10.6 Inferencing—Within Texts (Modifier)

Inferencing—Within Texts should be coded for activities that involve making inferences within a text based on information that has not been explicitly stated in the text but is inferred from information already conveyed in the text. An example of this would be if the students were reading a story about a boy who lost his dog, and the teacher asks the students, "How do you think the boy felt when he finally found his dog at the end of the story?"

7.1.10.7 Inferencing—Background Knowledge (Modifier)

Inferencing—Background Knowledge should be coded for activities that involve making inferences within a text based on information that has not been explicitly stated in the text but is based on activating student's background knowledge to make connections between their own knowledge/experiences and information presented in the text to make inferences about the story. An example of this would be if the teacher is reading a story about a boy who loses his dog, and the teacher asks the students, "Have any of you ever lost a pet? How did it make you feel? How do you think the boy in the story feels?" *The difference between Inferencing—Background Knowledge versus Prior Knowledge is that the teacher must* explicitly *ask the students to make an inference by activating background knowledge.*

7.1.10.8 Questioning (Modifier)

Listening and Reading Comprehension > Questioning should be coded for activities that involve generating or answering questions regarding factual or contextual knowledge from the text (e.g., "What did Ira miss when he went to the sleepover?" "What was the name of ______?"), provided that these activities are not better coded as Comprehension > Prior Knowledge (e.g., when the teacher uses a question to scaffold children in activating personal knowledge related to the text: "When you go to an amusement park, what do you expect to see?"), Comprehension > Monitoring (e.g., when the teacher uses a question aimed at stimulating students' metacognitive assessment of whether they comprehended the text: "Did I understand what happened there?"), or Comprehension > Predicting (e.g., when the teacher asks students to predict what will happen next: "What do you think the lost boy will do now?"). Questioning should also be coded for Accelerated Reader tests, which are typically completed on the computer; an Accelerated Reader test should also be coded as event code > Assessment. This code should also be used as a default code for activities when it is not clear whether the activity is highlighting, questioning, or summarizing.

7.1.10.9 Monitoring (Modifier)

Listening and Reading Comprehension > Monitoring should be coded for activities that involve stimulating students' metacognitive awareness regarding their comprehension of text, or sharing strategies to provoke students to think about whether they are fully understanding. Generally, these activities involve thinking about one's own understanding of a particular text and whether the text is making sense (e.g., the teacher pauses and says, "Did that make sense to you? If not, how can we fix it?" or "Wait, did I understand that?" or "That didn't make sense to me. Let's go back and reread"). These may include identifying areas of difficulty while reading, using think-aloud procedures to pinpoint difficulties, looking back in the text, restating or rephrasing text, or looking forward to solve a problem (last sentence from Pathways code). This may also involve the use of clarifying. For example, the teacher tells students, "As you read, make a note of sections that you do not understand and reread them to better understand what they say."

7.1.10.10 Highlighting/Identifying (Modifier)

Listening and Reading Comprehension > Highlighting/Identifying should be coded for activities that involve picking out the important details conveyed through a text. Examples include verbally listing, underlining, highlighting, or otherwise noting major points. Comprehension > Highlighting differs from Comprehension > Summarizing, because it explicitly involves identifying the important details within the text. This code can also be used if the teacher asks a student to name his or her favorite part of the story, and the student names important details as his or her favorite part; however, if the student does not name important details, or you cannot hear the student's answer, then code Listening and Reading Comprehension > MF-TBD. If it is not clear whether the activity is highlighting, questioning, or summarizing, code the activity as Listening and Reading Comprehension > Questioning.

7.1.10.11 Summarizing (Modifier)

Listening and Reading Comprehension > Summarizing should be coded for activities that involve generating an overall statement or identifying the main ideas of the content of the text. This activity should condense the text to the main points, which is much different and shorter than retelling. This could also include drawing a picture in response to the text just read. If it is unclear whether the activity is highlighting, questioning, or summarizing, then code the activity as Listening and Reading Comprehension > Questioning.

7.1.10.12 Context Cues (Modifier)

Listening and Reading Comprehension > Context Cues should be coded for activities in which students are using pictures, the title, or previous parts of the text to understand a new event or new information presented in the text. For example, a teacher might advise a student to look at a picture to identify the setting of a story.

7.1.10.13 Graphic/Semantic Organizers (Modifier)

Listening and Reading Comprehension > Graphic/Semantic Organizers should be coded for activities in which students are using graphic or semantic organizers (e.g., Venn diagrams, story webs) to aid their comprehension. Graphic/semantic organizers used to plan writing instruction should be coded under Writing > Prewriting/Organizers.

7.1.10.14 Prior Knowledge (Modifier)

Listening and Reading Comprehension > Prior Knowledge should be coded for activities that involve activating *students'* personal knowledge as it relates to the content of the text to facilitate comprehension. An example would be asking, "Have you ever slept over at a friend's house?" when reading *Ira Sleeps Over* by Bernard Waber. This relates to the student's personal knowledge. *If the teacher asks students to make an inference by using prior knowledge, then this would be coded as Inferencing—Background Knowledge.*

7.1.10.15 Retelling (Modifier)

Listening and Reading Comprehension > Retelling should be coded when students are asked to retell a story using their own words. This differs from Comprehension > Summarizing, because a retell should mimic the text structure and include *as many details of a text as possible.*

7.1.10.16 Sequencing (Modifier)

Listening and Reading Comprehension > Sequencing should be coded for activities that involve putting events from a text into the correct order. *If the activity involves graphic organizers, this should be noted in the Comments field.*

7.1.10.17 Comparing/Contrasting (Modifier)

Listening and Reading Comprehension > Comparing/Contrasting should be coded for activities that involve comparisons across or within texts. If the Comparing/Contrasting activity involves the use of a graphic organizer (e.g., Venn diagram), this should be coded under Comprehension > Graphic/Semantic Organizers with "compare/contrast" noted in the Comments field.

7.1.10.18 Cause and Effect (Modifier)

Listening and Reading Comprehension > Cause and Effect should be coded when the teacher and/or students are discussing cause and effect. An example of this would be if the teacher says, "There was a rock in the middle of the sidewalk. The boy fell down. What would be the cause, and what is the effect?" The teacher/activity should explicitly state that students are to give the cause and effect. This activity may also involve the discussion of the concept of cause and effect.

7.1.10.19 Fact vs. Opinion (Modifier)

Listening and Reading Comprehension > Fact vs. Opinion should be coded when the teacher and/or students are discussing opinion versus fact. An example of this would be if the teacher gives an example from a text (e.g., "The balloon in the story was red; is this an opinion or a fact?"). The teacher/activity should explicitly state that the students are to decide whether a particular sentence or part of the text is an opinion or a fact. This should not be confused with the Inferring codes when the purpose is to make inferences.

7.1.10.20 Multicomponent/Integrated Comprehension Strategy (Modifier)

The Multicomponent/Integrated Comprehension Strategy code should be used when students are being taught to use a combination of comprehension strategies to comprehend a text. This code can be used receptively and/or expressively, includes the intentional or systematic use of combining strategies, and prompts metacognition. An example of this activity would be if the students are being taught to use the UNRAAVEL strategy (e.g., systematically underlining keywords, numbering paragraphs) or reciprocal teaching. For example, students are given a passage and are told to identify keywords before reading it, so they can find the words more easily later on. Another example would be the teacher asking, "What kinds of comprehension strategies might we use to understand this passage?" or "Where would we find more information about this topic?" In the latter cases, the students are implicitly asked to select among numerous comprehension strategies.

7.1.10.21 MF-TBD

The Listening and Reading Comprehension > MF-TBD code should be used only when (a) none of the other Comprehension modifier codes are appropriate for a given activity, and (b) the activity fits the Comprehension description. *A brief description of the activity should be noted in the Comments field.* Note that by definition, these activities should be meaning-focused. If the teacher asks a student to name his or her favorite part of the story, and the child does not name important details from the story, or you cannot hear the student's answer, then code the activity as Listening and Reading Comprehension > MF-TBD. If the student does give important details, then code as Listening and Reading Comprehension > Highlighting.

Appendix B

Quality Ratings Rubric Descriptions

Classroom implementation of individualized instruction	Classroom orienting, organization, and planning	Robust vocabulary instruction	Warmth and responsiveness, control, and discipline
Teacher fidelity rating 1			
The teacher is not differentiating instruction.	The classroom is not organized. Transitions are long, and instructional delivery is unclear and confusing. The general feeling is of chaos.	The teacher is not providing any vocabulary instruction.	The teacher is detached or overcontrolling and often punitive (i.e., neglectful, authoritarian), is nonresponsive, and does not select or incorporate students' responses, ideas, examples, and experiences into the lesson.
Teacher fidelity rating 2			
The teacher uses primarily whole-class instruction. When small groups are used, they are not always focused on literacy. Instructional delivery is inconsistently paced for students with varying skill levels.	The classroom has inconsistent organization. Transitions are of long to reasonable duration and are inefficient. Limited instructional clarity (e.g., the teacher's instructions to students regarding how to complete activities are not always easy for students to understand).	The teacher provides some vocabulary instruction, but it is largely by defining words and, sometimes, using words in sentences. There are no opportunities for using the vocabulary in other contexts. Words are frequently Tier 1, with few Tier 2 or Tier 3 words selected.	The teacher is somewhat detached or overcontrolling and fairly punitive (i.e., authoritarian, neglectful) or indulgent. Whenever discipline is imposed, it is inconsistent and only occasionally effective. The teacher is rarely responsive and rarely selects and incorporates students' responses, ideas, examples, and experiences into the lesson.
Teacher fidelity rating 3			
There is clear evidence of differentiation. The teacher uses small groups; however, the students in the small groups generally receive highly similar amounts and types of instruction.	The classroom is reasonably organized, instructional clarity is evident, and transitions are fairly efficient.	The teacher provides fairly adequate vocabulary instruction, which may extend beyond simple definitions occasionally, and provides some opportunities for using the words in other contexts. There is some attempt to be intentional about selecting Tier 2 words, but only about one third of the words contribute meaningfully to students' understanding.	The teacher is occasionally detached or overcontrolling. The teacher is fairly responsive and effective at selecting and incorporating students' responses, ideas, examples, and experiences into the lesson.
Teacher fidelity rating 4			
There is clear evidence of differentiation. The teacher uses small groups, and there is evidence that instruction is individualized. Most of the language arts block is spent in meaningful literacy activities.	The classroom is fairly well organized, and there is adequate but not excellent instructional clarity. Instruction is usually planned in advance.	The teacher provides adequate vocabulary instruction, which may extend beyond simple definitions, and provides some opportunities for using the words in other contexts. There are attempts to be intentional about selecting Tier 2 words, but only about half of the words contribute meaningfully to students' understanding and are not always relevant to the text.	The teacher is authoritative and interacts with students. The teacher is responsive and usually effective at selecting and incorporating students' responses, ideas, examples, and experiences into the lesson. The classroom tends to offer a positive learning environment with clear expectations for students as members of the classroom learning community.

(*continued*)

Classroom implementation of individualized instruction	Classroom orienting, organization, and planning	Robust vocabulary instruction	Warmth and responsiveness, control, and discipline
Teacher fidelity rating 5			
The teacher uses small groups, and there is good evidence that the instruction is individualized. The number and composition of groups are based on effective group size and the range of literacy skills of the students.	The classroom is well organized, and there is good instructional clarity. Transitions are efficient, and instructional delivery is well paced. Most of the language arts block is spent in meaningful literacy activities.	The teacher provides good vocabulary instruction, which typically extends beyond simple definitions, and provides students with opportunities for using the words in other contexts, including written and media contexts. Selecting Tier 2 words is intentional, and words contribute meaningfully to students' understanding of relevant text.	The teacher is authoritative, responsive, interactive, and usually effective at selecting and incorporating students' responses, ideas, examples, and experiences into the lesson. The teacher is usually effective at securing and maintaining students' attention as needed. The classroom usually offers a positive learning environment with clear expectations for students as members of the classroom learning community.
Teacher fidelity rating 6			
Teachers who fully implement Individualizing Student Instruction use multiple and flexible student grouping configurations and regrouping of students based on formal or informal assessment data. The content of literacy instruction is differentiated.	The classroom and instruction are well organized. Classroom routine is evident. Transitions are efficient. The entire language arts block is spent in meaningful literacy activities.	The teacher provides exemplary vocabulary instruction, which always extends beyond simple definitions, and provides students with opportunities for using words in other contexts, including written and media, at other times of the day or week, or at home. Selecting Tier 2 words is intentional, and the words contribute meaningfully to students' understanding of relevant text. Tier 3 words used are appropriate.	The teacher is authoritative and highly responsive to students, and the classroom is highly interactive, with all students participating in learning. The teacher effectively selects and incorporates students' responses, ideas, examples, and experiences into the lesson. The classroom consistently offers a positive learning environment with clear expectations for students as members of the classroom learning community.

Note. For a 1 rating, the teacher is consistently weak in this area. For a 3 rating, the teacher shows the characteristic but is inconsistent. For a 5 rating, the teacher is consistently strong in this area. For a 6 rating, the teacher is exemplary. Indicators and checklist are available upon request from the first author.

Appendix C

Sample Coding Output of Third-Grade Classroom Instruction

Coding Output for Literacy Block of ISI Treatment Teacher Observed in May 2009

Instruction coding was completed for 10 randomly selected target students. The teacher began the literacy block by describing how to complete the individual and center activities that the students were to complete while she worked with each small group of students. Groups generally included four students and followed the A2i grouping recommendations. To facilitate reading the output, we have included coding only for students B and L. Student B is a multiracial boy who achieved a passage comprehension GE of 2.9 in the fall and 3.8 in the spring. Student L is an African American girl who achieved a GE of 3.8 in the fall and 4.8 in the spring. Seventy-two percent of the students at the school qualified for the school's FARL.

Time represents time elapsed in seconds from the beginning of the literacy block to the initiation of a new activity. Any activity that lasted at least 15 seconds was coded. Please see Appendix A for excerpts from the coding manual. The entire manual is available upon request from the first author. All content codes following the context and management codes represent the content of the activities provided within that context.

Time on the videotape (seconds)	Student	Context, management, and content	Type of content	Materials	Comments
0	B	Language arts			
0	L	Language arts			
0	B	Organization, whole class, TSM			
0	L	Organization, whole class, TSM			
0	B	Organization	Orient day		
0	L	Organization	Orient day		
39.941	B	Organization	Orient class		
39.941	L	Organization	Orient class		
57.956	B	Noninstruction, whole class, TSM			
57.956	L	Noninstruction, whole class, TSM			
57.956	B	Noninstruction	Waiting		
57.956	L	Noninstruction	Waiting		
74.633	B	Organization, whole class, TSM			
74.633	L	Organization, whole class, TSM			
74.633	B	Organization	Orient day		
74.633	L	Organization	Orient day		
91.917	B	Noninstruction, whole class, TSM			
91.917	L	Noninstruction, whole class, TSM			
91.917	B	Noninstruction	Waiting		
91.917	L	Noninstruction	Waiting		
108.488	B	Organization, whole class, TSM			
108.488	L	Organization, whole class, TSM			
108.488	B	Organization	Orient activity		
108.488	L	Organization	Orient activity		
124.76	B	Noninstruction, individual, SM			
124.76	L	Noninstruction, individual, SM			
124.76	B	Noninstruction	Transition/act		
124.76	L	Noninstruction	Transition/act		

(*continued*)

Time on the videotape (seconds)	Student	Context, management, and content	Type of content	Materials	Comments
144.952	B	Noninstruction, small group, peer-managed			With peers
145.327	L	Noninstruction, small group, TSM			
145.327	L	Noninstruction	Waiting		
175.053	L	Listening and reading comprehension	Prior knowledge	Expository text—science	
175.053	L	Small group, TSM-MF			
190.343	B	Noninstruction, individual, SM			
214.825	L	Organization, small group, TSM			
214.825	L	Organization	Orient activity		
224.982	B	Individual, SM-MF			
224.983	B	Text reading	Silent sustained reading	Workbook/worksheet—science	
237.799	L	Small group, TSM-MF			
237.799	L	Text reading	Student read-aloud/ Individual	Expository text—science	
554.917	B	Noninstruction, individual, SM			
554.918	B	Noninstruction	Transition/act		
564.909	L	Text reading	Student read-aloud/ individual	Expository text—science	
580.066	L	Listening and reading comprehension	Schema building	Expository text—science	
601.924	L	Listening and reading comprehension	Questioning	Expository text—science	
604.689	B	Individual, SM-MF			
604.69	B	Listening and reading comprehension	Questioning	Workbook/worksheet—science	
626.082	L	Listening and reading comprehension	Schema building	Expository text—science	
648.173	L	Noninstruction, small group, TSM			
648.173	L	Noninstruction	Waiting		
649.516	B	Noninstruction, individual, SM			
649.516	B	Noninstruction	Transition/act		
669.953	B	Individual, SM-MF			
669.954	B	Listening and reading comprehension	Questioning	Workbook/worksheet—science	
670.231	L	Listening and reading comprehension	Inferencing—background knowledge	Expository text—science	
699.875	L	Text reading	Student read-aloud/ individual	Expository text—science	
770.034	L	Listening and reading comprehension	Questioning	Expository text—science	
779.823	B	Noninstruction, individual, SM			
779.825	B	Noninstruction	Transition/act		
795.125	L	Organization, small group, TSM			
795.125	L	Organization	Orient class		
820.211	L	Small group, TSM-MF			
820.228	L	Text reading	Student read-aloud/ individual	Expository text—science	
824.992	B	Individual, SM-MF			
824.996	B	Listening and reading comprehension	Questioning	Workbook/worksheet—science	
884.795	L	Listening and reading comprehension	Summarizing	Expository text—science	

(*continued*)

Appendix C (*continued*)

Time on the videotape (seconds)	Student	Context, management, and content	Type of content	Materials	Comments
914.698	L	Noninstruction, small group, TSM			
914.698	L	Noninstruction	Waiting		
950.412	L	Small group, TSM-MF			
950.412	L	Text reading	Student read-aloud/ individual	Expository text—science	
989.525	B	Noninstruction, individual, SM			
989.53	B	Noninstruction	Transition/act		
1,019.984	B	Individual, SM-MF			
1,019.987	B	Listening and reading comprehension	Questioning	Workbook/worksheet—science	
1,025.044	L	Listening and reading comprehension	Questioning	Expository text—science	
1,084.991	L	Text reading	Student read-aloud/ individual	Expository text—science	
1,144.968	L	Listening and reading comprehension	Schema building	Expository text—science	
1,184.935	B	Noninstruction, individual, SM			
1,184.936	B	Noninstruction	Transition/act		
1,195.113	L	Text reading	Student read-aloud/ individual	Expository text—science	
1,214.896	L	Small group, TSM-CF			
1,215.073	L	Phonological awareness	Syllable counting		Counting syllables
1,234.899	L	Small group, TSM-MF			
1,235.153	L	Text reading	Student read-aloud/ individual	Expository text—science	
1,270.454	B	Individual, SM-MF			
1,270.455	B	Text reading	Silent sustained reading	Workbook/worksheet—language arts	
1,355.147	L	Listening and reading comprehension	Inferencing—background knowledge	Expository text—science	
1,389.605	L	Text reading	Student read-aloud/ individual	Expository text—science	
1,406.44	L	Text reading	Teacher read-aloud/ students listening	Expository text—science	
1,436.424	L	Text reading	Student read-aloud/ individual	Expository text—science	

Note. CF = code-focused. MF = meaning-focused. SM = student-managed. TSM = teacher/student-managed.