



Published in final edited form as:

*J Educ Psychol.* 2014 August ; 106(3): 762–778. doi:10.1037/a0035921.

## **Capturing the complexity: Content, type, and amount of instruction and quality of the classroom learning environment synergistically predict third graders' vocabulary and reading comprehension outcomes**

**Carol McDonald Connor,**

Arizona State University and the Learning Sciences Institute

**Mercedes Spencer,**

Florida State University and the Florida Center for Reading Research

**Stephanie L. Day,**

Arizona State University and the Learning Sciences Institute

**Sarah Giuliani,**

Florida State University

**Sarah W. Ingebrand,**

Arizona State University and the Learning Sciences Institute

**Leigh McLean, and**

Arizona State University and the Learning Sciences Institute

**Frederick J. Morrison**

University of Michigan

### **Abstract**

We examined classrooms as complex systems that affect students' literacy learning through interacting effects of content and amount of time individual students spent in literacy instruction along with the global quality of the classroom-learning environment. We observed 27 third grade classrooms serving 315 target students using two different observation systems. The first assessed instruction at a more micro-level; specifically, the amount of time individual students spent in literacy instruction defined by the type of instruction, role of the teacher, and content. The second assessed the quality of the classroom-learning environment at a more macro level focusing on classroom organization, teacher responsiveness, and support for vocabulary and language. Results revealed that both global quality of the classroom learning environment and time individual students spent in specific types of literacy instruction covering specific content interacted to predict students' comprehension and vocabulary gains whereas neither system alone did. These findings support a dynamic systems model of how individual children learn in the context of classroom literacy instruction and the classroom-learning environment, which can help to improve observations systems, advance research, elevate teacher evaluation and professional development, and enhance student achievement.

---

Reading comprehension and vocabulary have been identified as strong predictors of future academic success (NICHD, 2000) as well as of overall school and life outcomes (Beck, McKeown, & Kucan, 2002). Yet, by the end of 4th grade only about 34% of U.S. students are reading and comprehending proficiently (National Center for Education Statistics, 2013). Accumulating research points to the importance of classroom literacy instruction and the opportunities to learn that students receive in the early grades (Connor et al., 2013; NICHD, 2000; Pianta, Belsky, Houts, Morrison, & NICHD-ECCRN, 2007; Snow, 2001; Tuyay, Jennings, & Dixon, 1995). Understanding the classroom learning environment is important and finding ways to elucidate the active ingredients of this environment that predict student outcomes are essential but challenging.

In this study, we used a dynamic systems framework (Yoshikawa & Hsueh, 2001), which holds that there are multiple sources of influence on children's learning (Bronfenbrenner & Morris, 2006) including the instruction they receive, how this instruction is delivered (Connor, Piasta, et al., 2009; Reis, McCoach, Little, Muller, & Kaniskan, 2011), the general climate of the classroom (Rimm-Kaufman, Paro, Downer, & Pianta, 2005), teacher characteristics (Raver, Blair, & Li-Grining, 2011), and students themselves (Connor & Morrison, 2012; Justice, Petscher, Schatschneider, & Mashburn, 2011). Further, these sources of influence interact in different ways with some seemingly important factors (e.g., teacher education) having relatively small effects on students' reading development (Goldhaber & Anthony, 2003) and other factors (e.g., content and minutes of instruction) having large effects (Connor, Morrison, Schatschneider, et al., 2011). High quality literacy instruction should provide students with individualized opportunities to learn that, in turn, influence their reading comprehension and language development (Beck et al., 2002; Beck, Perfetti, & McKeown, 1982; Connor et al., 2013; Snow, 2001). Thus, there is an increasing policy and research focus on how to measure classroom instruction in ways that validly and robustly predict gains in students' literacy and vocabulary skills (see Crawford, Zucker, Williams, Bhavsar, & Landry, in press; Kane, Staiger, & McCaffrey, 2012; Ramey & Ramey, 2006; Reddy, Fabiano, Dudek, & Hsu, in press; Whitehurst et al., 1988). The aim of this study is to systematically investigate the classroom learning environment as a dynamic system, identify major dimensions of classroom instruction -- at both the individual student level and the global classroom level -- that may influence students' literacy achievement and how these dimensions might work together synergistically to support (or fail to support) opportunities for learning that result in gains in third graders' vocabulary and reading comprehension.

## Classroom Observation Systems

Teacher value-added scores have revealed that there is measurable variability in the effectiveness of teaching, which has direct implications for students' success or failure (Konstantopoulos & Chung, 2011). However, value-added scores do not reveal what is going on in the classroom and the characteristics of the environment that explain the variability in teachers' value-added scores. The development of rigorous classroom observation systems that are reliable and have good predictive validity are important because they help to open up the black box of classroom instruction, so to speak, and begin to move us toward what has been described as "shared instructional regimes" (Raudenbush,

2009). Raudenbush describes historical and recent theories of teaching as “privatized idiosyncratic practice” (p. 172) whereby teachers close their classroom doors and teach in the ways they believe to be best and where the ideal teacher develops his or her own curriculum. The “idiosyncratic” practice of teachers who have a good grasp of the current research, who have expert and specialized knowledge of their content area, and who understand how to use research evidence to inform their practice can be highly effective. However, the privatized idiosyncratic practice of some teachers may be highly ineffective (Piasta, Connor, Fishman, & Morrison, 2009), particularly for children from low SES families whose home learning environment and access to resources is limited and who are more reliant on the instruction they receive at school. Research-based observation tools allow us to illustrate what effective expert practice in the classroom actually looks like so that it can be shared among a community of professionals – both educators and researchers – to improve teaching.

There are several well-documented observation systems in use with new systems being developed (Connor, 2013a). These classroom observation systems provide important insights and most of them explain at least modest amounts of the variance in students’ literacy learning. For example, Kane and colleagues (2012) tested several observation instruments, including the Framework for Teaching (FFT; Danielson, 2007), CLASS (Pianta et al., 2008), Protocol for Language Arts Teaching (PLATO, Grossman et al., 2010), Mathematical Quality of Instruction (MQI; Hill, Ball, Bass, & Schilling, 2006), and UTeach Teacher Observation Protocol (UTOP, 2009). Results revealed that although none of the systems designed to assess English/Language Arts instruction correlated with teacher value-added scores computed using state-mandated assessments of English/Language Arts, they were mildly to moderately positively correlated with teacher value-added scores computed using the SAT-9 reading assessment.

### **Classroom Observations in the Present Study**

We used two different observation coding systems to test the dynamic systems model of instruction in the current study: the quality of the classroom learning environment (CLE) and ISI/Pathways-observation system (ISI/Pathways, Connor, Morrison, et al., 2009). The first was designed to capture the global quality of the CLE using a rubric that captured elements of the CLE that are generally predictive of student outcomes. The second, ISI/Pathways, was designed to record the amount of time individual students spent in various types of literacy instruction; the content of this instruction; the role of the teacher; and the context (e.g., whole class, small group) in which instruction was provided. We conjectured, following the dynamic systems model, that classroom opportunities to learn would operate at both student and classroom levels and that the two systems together might better elucidate the complexities of the classroom and effective learning opportunities afforded to students than either system alone. We describe each below.

**Quality of the Classroom Learning Environment Rating Scale**—The CLE rating scale (see Appendix A) was designed to rate the classroom on three dimensions: Teacher Warmth, Responsiveness and Discipline; Classroom Organization; and Teacher Support for Vocabulary and Language Development – with one rating for each scale for the entire

observation of the literacy block. *Teacher warmth, responsiveness, and discipline* were defined as teachers' regard for their students, the overall emotional climate of the classroom as well as the way in which they responded to students, particularly with regard to how they responded to student misbehavior and disruptions (Pianta, La Paro, Payne, Cox, & Bradley, 2002). Examples of teacher warmth and responsiveness include being supportive of students, providing positive feedback, clearly communicating what is expected of students, and providing discipline in a positive and supportive way (Rimm-Kaufman et al., 2005). The kinds of discussions and types of questions used, for example, coaching versus telling (Taylor & Pearson, 2002), were measured indirectly through this dimension. Research has shown that students whose teachers were more warm and responsive achieved greater gains in reading skills, including vocabulary, by the end of first grade (Connor, Son, Hindman, & Morrison, 2005).

*Classroom organization* is defined as the degree to which the teacher takes time to give students thorough directions for upcoming activities, has clear rules for behavior, and has established routines that optimize student learning time (Wharton-McDonald, Pressley, & Hampston, 1998). When teachers have strong orienting and organizational skills, they are better able to create an efficient and productive CLE. It has been found that teachers who implement rules and effectively establish routines are less likely to have difficulties with classroom management (Borko & Niles, 1987; Cameron, Connor, Morrison, & Jewkes, 2008).

According to Beck and colleagues (2002), teacher support for vocabulary and language development should be "robust," meaning that instruction should include activities beyond those that encourage rote memorization of words and their definitions and, instead, involve rich contexts that extend beyond the classroom. Such support has the potential to improve language skills overall given that processes where vocabulary knowledge is highly utilized (e.g., during reading comprehension) require skills above and beyond knowing the definitions of words. Therefore, instructional techniques that take into consideration that vocabulary knowledge is part of students' background knowledge (Stahl, 1999), rather than a singular component, are more likely to be effective. Such techniques encourage students to actively use and think about word meanings and create word associations in multiple contexts. Accumulating evidence further highlights the importance of supporting student language development because vocabulary (and oral language skills in general) are highly predictive of students' reading comprehension (Biemiller & Boote, 2006; Cain, Oakhill, & Lemmon, 2004), and at the most basic level, allow students to understand grade-level texts. Yet, despite the known importance of vocabulary and language skills to later reading abilities and subsequent academic success, it has been shown that vocabulary instruction is often missing from language arts/literacy classrooms (see Cassidy & Cassidy, 2005/2006; Rupley, Logan, & Nichols, 1998).

**The Individualizing Student Instruction/Pathways Classroom Observation System (ISI/Pathways)**—The ISI/Pathways system (Connor, Morrison, et al., 2009) measures the amount of time (min;sec) individual students within a classroom spend in literacy instruction activities across three dimensions: content of instruction; context; and the role of the teacher and student in the learning activity. The *content of instruction dimension*

(see Table 1 and Appendix B) captures the specific topic of the literacy instruction that individual students are receiving (e.g., comprehension, vocabulary, text reading). We also coded non-instructional activities, which included off-task or disruptive behaviors, transitions between activities, or time spent when the teacher was giving directions for upcoming activities. The *context dimension* captures the student grouping arrangement and includes whole class, small group, or individual instruction. *Management* captures who is controlling the students' attention during an activity: the teacher and student working together (teacher-child managed), peer-managed (students working with each other), or child-self managed (student managing his/her own attention). These dimensions operate simultaneously (see Table 1) to describe instructional and non-instructional activities observed during reading instruction. For example, the teacher and students discussing a book they just read together would be coded as comprehension (meaning-focused) teacher/child-managed, whole class activity that lasted for 11 minutes.

A key characteristic of ISI/Pathways is that the measurement of instructional time and content is assessed for individual students in the classroom (Connor, Morrison, et al., 2009). Hence, the system is able to capture the learning opportunities afforded to each student; for example, recording that Student A was reading with the teacher while, at the same time, Student B was off task and not redirected. A global classroom-level system would likely capture Student A's instructional opportunities but not Student B's. The more precise measure of each individual student's learning opportunities has been used to identify instructional practices as well as child characteristic X instruction interaction effects on students' reading achievement (Connor, Morrison, Fishman, et al., 2011; Connor, Morrison, Schatschneider, et al., 2011; Connor, Piasta, et al., 2009). Across studies, amounts, content, context, and types of instruction measured by the ISI/Pathways Observation system predicted student literacy achievement (Connor, Morrison, et al., 2009), particularly the difference between observed and recommended individualized types/content and amounts of instruction (Connor, Piasta, et al., 2009). The closer the observed amount matched the recommended amount, the greater the students' literacy gains were.

We posed the following research question:

How does combining measures of the duration of different types of literacy instruction and content for individual students with a more global measure of the CLE synergistically affect students' reading comprehension and vocabulary outcomes? Using our dynamic systems model of the classroom, we hypothesized that neither the quality of the CLE nor the amount of time individual students spent in different types/content of literacy instruction (ISI/Pathways) would be strong independent predictors of vocabulary and comprehension outcomes for third grade students. Rather, we hypothesized that there would be interaction effects involving both systems that would significantly and positively predict third graders' language and comprehension gains. Such interaction effects would, hypothetically, better capture the complexity of classroom instruction and the learning environment.

## Methods

### Participants

The participants included third grade teachers (n=27, 13 in the individualized reading group and 14 in the vocabulary control group) and their students (n = 315) in 7 schools who were participating in a randomized controlled study evaluating the efficacy of individualized reading instruction from first through third grade (Connor, Morrison, Fishman, et al., 2011). We selected third grade because comprehension of text becomes increasingly important (Gottardo, Stanovich, & Siegel, 1996; Reynolds, Magnuson, & Ou, 2010) as children move from *learning to read* to *reading to learn* (Chall, 1967). The schools were located in an economically and ethnically diverse school system in north Florida.

**Teachers**—All teachers completed the study with the exception of three teachers who were not present during the last month of the study; however, results for these teachers and their students were used in the analysis because observations were completed before they left and all of their participating students were assessed. All of the teachers met the state certification requirements and had at least a bachelor's degree related to an educational field. Seven of the teachers had certifications or degrees beyond a bachelor's degree. Teachers' classroom teaching experience ranged from 0 to 30 years, with a mean of 10.9 years of experience.

Teachers in both conditions participated in half-day workshops in the fall and again in January for either literacy or vocabulary. They also participated in monthly meetings. Teachers in the reading intervention also received bi-weekly classroom-based support (not on the day observed). In total, teachers in the vocabulary condition received about 12 hours of professional development and teachers in the individualized reading intervention received between 18-20 hours (some needed more help with the technology). Professional development for the individualized reading intervention helped teachers learn how to individualize student instruction and how to be better organized. Professional development for the vocabulary intervention focused on vocabulary teaching methods described in Beck et al., (2002). Results of this study revealed that students whose teachers were in the individualized reading intervention group demonstrated significantly greater gains in reading comprehension than did the students whose teachers were in the vocabulary intervention group. Results are fully described in Connor et al., 2011.

**Students**—School-wide percentages of students qualifying for free and reduced lunch programs (FARL) ranged from 92% (high poverty) to 4% (affluent). All schools used the Open Court Reading Curriculum and had a 90-minute uninterrupted block of time devoted to reading instruction. All observations were conducted during this literacy block.

Student participant demographics were collected through parent reports and school records and were as follows: 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% indicated other ethnic groups. Forty-seven percent qualified for FARL. Approximately 12% qualified for special education services. A subset of students from each classroom was randomly selected to be the focus of observation coding using the following procedure: Because this was a first through third grade longitudinal study (Connor et al., 2013), third

graders who were in the first and second grade studies were automatically selected as target students. We then randomly selected from among their classmates to bring the total number of target students to a minimum of 8 per classroom after rank ordering and randomly selecting within terciles so that we had a distribution of reading skill level. This provided the final sample of 315 students. On average, there were 11 target students per classroom and this ranged from 6 to 19. Three classrooms had 6 target students; two had 7 students; all others had 8 or more target students. Comparisons of this sub-sample with the entire sample revealed no significant differences on any of the measures of interest.

### **Observation of Instruction and Classroom Learning Environment**

Again, in this study we used two different observation systems – the ISI/Pathways system, which measured the amounts and types/content of instruction students received and the CLE quality rubric, which captured the quality of the classroom-learning environment. Both systems used the same videotaped classroom observations. Classrooms were videotaped three times, once during the fall, once in winter, and once in the early spring of the academic year. This video footage of the 90-minute block of time devoted to literacy instruction was captured using two digital camcorders with wide-angle lenses. Cameras were not focused on individual students per se. Rather, cameras were positioned at opposite sides of the classroom to capture as much of the class as possible. However, during small group instruction, one camera was focused on the teacher's small group while the second camera captured students working independently and the other small groups. While video-recording, trained research assistants kept 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). These notes were used in conjunction with video footage during coding and provided information for coders about students or activities that could not easily be seen on the videos. Observations were scheduled at the teachers' convenience so the assumption was that the instruction was of the highest quality the teacher could provide.

**Quality of the Classroom Learning Environment**—The quality of the CLE was assessed using a detailed rubric/rating scale (see Appendix A). The rating scale ranged from 1 (low) to 6 (high) and examined three global classroom-level dimensions – organization, support for vocabulary and language, and teacher responsiveness – with priority given to specific aspects of the CLE that were the focus of the professional development provided and that previous research has associated with more effective instruction (i.e., higher quality) (Brophy, 1979; Cameron, Connor, & Morrison, 2005; Pianta et al., 2002; Snow, Burns, & Griffin, 1998; Taylor, Pearson, Clark, & Walpole, 2000; Wharton-McDonald et al., 1998). Trained research assistants coded all three videotaped observations using the CLE rubric. Highly trained research assistants who were blind to the teachers' treatment assignment coded the video footage. Sufficient inter-rater reliability on fall observations, based on Landis and Koch (1977) criteria, was reached prior to coding the winter observation (Cohen's Kappa = 0.73). Approximately 10% of coded winter and spring footage was randomly selected and re-coded, and inter-rater reliability was maintained (Cohen's Kappa = 0.73). The winter observation CLE score was used in this study because teaching tends to be more consistent during the winter months (Hamre, Pianta, Downer, &

Mashburn, 2007) than in the earlier months, when routines are getting established and teachers are just getting to know their students. Two teachers' scores were based on the spring observation because a student teacher, and not the primary teacher, was teaching during the winter observation. Because scores on the three scales were moderately correlated ( $r = .59-.60$ ) and combining the 3 scores improved internal reliability, the scores were summed to provide a total CLE score.

**ISI/Pathways Observation System**—As noted previously, the ISI/Pathways system was designed to document instruction across three dimensions: (1) the *content* of the literacy instruction (e.g., comprehension, oral language); (2) the *context* (i.e., small group, individual or whole class); (3) the extent to which the teacher was interacting with students (*management*) which included teacher/child-managed instruction (teacher and students working together) or child-managed instruction (students working with each other or independently) (Connor, Morrison, et al., 2009). Using Noldus Observer Video-Pro software (XT version 8.0), instructional activities observed during the 90-minute literacy block, which lasted 15 seconds or longer, were coded for each target student. Non-instructional practices (including transitions) were also coded. An excerpt from the ISI/Pathways Coding Manual is presented in Appendix B.

Trained research assistants coded each of the videos using the Noldus Observer XT software. The training process was extensive, typically lasting 4-6 weeks until coders achieved adequate inter-rater reliability (Cohen's Kappa > 0.7) with a master coder. Questions about coding were discussed at bi-weekly coding meetings until consensus was achieved. Random selection and analysis of approximately 10% of the videos revealed good ongoing inter-rater reliability among the coders (mean Cohen's Kappa = 0.72; Landis & Koch, 1977). The lengths of videos varied slightly depending on the season. Fall mean observation length was 85 minutes (standard deviation [SD] = 30), 73 minutes in the winter (SD = 35), and 79 minutes in the spring (SD = 23).

The observation system provided a detailed description of over 200 instructional variables. For this study, we separated type of instruction into code-focused and meaning-focused instruction (see Table 1). Code-focused instruction consisted of five distinct types: phonological awareness, morpheme awareness, word decoding, word encoding, and fluency. Meaning-focused instruction consisted of six distinct types: print and text concepts, oral language and oral vocabulary, print vocabulary, listening and reading comprehension, text reading, and writing. These distinctions were made using our theoretical framework that the largely unconscious and more automatic processes involved in reading, including sub-sentence processes (e.g., morphological awareness), and those that supported fluency (i.e., automaticity) were considered code-focused, whereas more reflective and text-level processes were meaning-focused (Connor, 2013b). The case can be made that morphological awareness should be considered a meaning-focused activity (Carlisle, 2000). Unfortunately, too little morphological awareness instruction was observed to test this alternative. For each student, the time (in seconds) spent in particular types of instruction (e.g., teacher/child-managed small group meaning-focused) were computed for fall, winter, and spring and then aggregated by taking the mean amount for each student.



## Student Assessments

Trained research assistants assessed students' language and literacy skills in the fall and again in the spring. Two measures of comprehension and two of vocabulary were utilized in this study. Comprehension was assessed using the Woodcock-Johnson-III Passage Comprehension subtest (WJ-III, Woodcock, McGrew, & Mather, 2001) and level 3 Gates–MacGinitie Reading Tests (GMRTs; MacGinitie & MacGinitie, 2006) Reading Comprehension subtest. Vocabulary was assessed using the WJ-III Picture Vocabulary subtest and the GMRT Reading Vocabulary subtest. Alternate versions of the assessments were administered in the fall and spring. Scores were provided to teachers and parents.

The WJ-III Passage Comprehension uses a cloze procedure where students read a sentence or short passage and supply the missing word. The Picture Vocabulary task asks students to name increasingly unfamiliar pictures. These subtests are administered individually and have demonstrated alpha reliability estimates of .88 for passage comprehension and .81 for picture vocabulary. W scores were used in the analyses because they are on an equal-interval scale, similar to the Rasch score (Rasch, 2001).

In the GMRT Reading Comprehension assessment, students read passages of varying length and complexity based on excerpts from narrative and expository texts commonly used in schools. Students are then asked to answer multiple choice questions, including some that require fairly high-level inferencing. As the student progresses through the test, the text becomes more difficult and the questions demand more inferencing. The GMRT multiple-choice vocabulary assessment requires students to read and then choose the correct meaning of an underlined word within a short phrase from four possible answers. Alternate form reliability for the GMRT has been reported to range from .74 to .92, and test–re-test reliability has been reported as ranging from .88 to .92. Extended scale scores were used for data analyses, which have an equal interval scale.

## Analytic Strategies

**Structural Equation Modeling**—We used structural equation modeling to examine the constructs of language and comprehension, and we hypothesized that the language and literacy measures would comprise one latent variable (e.g., Mehta, Foorman, Branum-Martin, & Taylor, 2005), keeping in mind that the GMRT vocabulary assessment required the students to read. As anticipated, the four measures were correlated (see Table 2). We used confirmatory factor analyses (AMOS version 21) to compare one- and two-factor models (Kline, 1998). We provide results for the spring models in detail. Results were highly similar for the fall models. The two-factor model, (comprehension and vocabulary) provided only moderately acceptable fit (TLI = .939; CFI = .994; RMSEA = .110; AIC = 32.401), whereas the one factor model provided a superior fit (TLI = .969; CFI = .994; RMSEA = .079; AIC = 31.525). We created fall and spring factor scores using Principal Component Factor Analysis, which explained 72.8% of the variance in the fall scores and 71.3% of the variance in the spring scores. Loadings are provided in Table 2. The factor variable, Vocabulary/Comprehension (VocComp; z-score with a mean of 0 and an SD of 1), was used in the analyses. Such factor or latent variable scores have the advantage of reducing measurement error and better capturing the complex construct of interest (Keenan,

Betjemann, & Olson, 2008; Kline, 1998; Mehta et al., 2005; Vellutino, Tunmer, Jaccard, Chen, & Scanlon, 2004).

**Hierarchical Linear Modeling**—Due to the nested structure of the data, students nested within classrooms and schools, hierarchical linear modeling (HLM, Raudenbush & Bryk, 2002) was used to answer our research question. HLM analyses accounted for shared variance within classrooms and schools, resulting in more accurate effect sizes and non-inflated standard errors (Raudenbush & Byrk, 2002). Model specification occurred in several steps. Initially, an unconditional model was created. Variance at the classroom level was divided by total variance (the summation of student- and classroom-level variance) to obtain the intra-class correlations (ICCs). ICC values represent the proportion of variance falling between classrooms. A three-level model with students nested within classrooms and classrooms nested within schools was created. This model indicated no significant between-school variance. Therefore, a simpler two-level model with the spring VocComp factor score as the outcome ( $Y_{ij}$ ) was created. Starting with the unconditional model, we first created a model with only the ISI/Pathways variables entered at the student level. Next, we created a model with only CLE quality entered at the classroom level; we then created a combined model and tested for cross level interaction effects (ISI/Pathways [student] X CLE quality [classroom]). We also tested for students' fall VocComp score X instruction interactions (models are available upon request).

## Results

Students in this sample entered third grade with vocabulary skills generally in line with grade level expectations based on standard scores ( $M=100$ ,  $SD= 15$ ), and percentile ranks ( $M=50$ ), which control for age. For example, their fall mean WJ Picture Vocabulary score was 98.92 and their spring mean score was 99.21 (see Table 2), suggesting grade/age appropriate gains from fall to spring, on average. Based on GMRT results, students were achieving slightly above grade expectations for both reading comprehension and reading vocabulary with percentiles in the fall of 51.81 and 55.04, respectively, and spring scores of 52.07 and 59.84, respectively. Only on the WJ Passage Comprehension measure did students generally score below test expectations with standard scores of 94.75 in the fall and 95.96 in the spring. Notably, students' standard scores and percentile ranks were the same or higher in the spring compared to fall, suggesting generally grade-appropriate gains in reading and vocabulary.

Overall, we observed high quality but variable CLE in these third grade classrooms with total scores ranging from 5 to 17 where a perfect score was 18 ( $M = 12.58$ ,  $SD = 3.14$ ). When considering the individual scales, teachers generally scored highest on the Orienting/Planning scale ( $M = 4.73$ ,  $SD = 1.09$ ) and lowest on the Support for Language scale ( $M = 3.66$ ,  $SD = 1.21$ ). They were rated a mean of 4.30 ( $SD = 1.34$ ) on the Responsiveness/Discipline scale.

An examination of the kinds of literacy activities that occurred during reading comprehension and vocabulary instruction using the ISI/Pathways observation system revealed variability in overall amounts and types of instruction (see Table 3, Figure 1, and

Appendix C). Beginning first with oral language and print vocabulary instruction, on average, students received about two minutes per day of small-group teacher/child-managed instruction and about three minutes per day of small-group child-managed instruction. Of this time, most was spent using vocabulary (Vocabulary Use, see Appendix B) and defining words with many of the child-managed activities using workbooks. See Figure 1 for graphs showing the amounts per day (seconds) of the various types of vocabulary activities observed. Examining teacher/child-managed small-group instruction in listening and reading comprehension revealed that students spent approximately four minutes per day in these activities with most of the time spent with the teacher asking questions and students responding (2.2 minutes). Generally, more than seven minutes per day were spent in whole-class teacher/child-managed vocabulary and reading comprehension activities with most of that time spent in question and answer time (3.1 minutes). In all cases, the ranges were large with some students receiving little vocabulary and comprehension instruction and some receiving much more.

Using these data, we created eight different variables of amounts and types/content of instruction (see Tables 1 and 3) that capture the entire duration of meaningful instruction that occurred during the literacy block: teacher/child-managed whole-class meaning-focused instruction; teacher/child-managed small-group meaning-focused instruction; teacher/child-managed whole-class code-focused instruction; teacher/child-managed small-group code-focused instruction; child/peer-managed whole-class meaning-focused; child/peer-managed small-group meaning-focused; child/peer-managed whole-class code-focused; and child/peer-managed small-group code-focused instruction. To create these variables, the data were first exported from Observer Pro and then cleaned and examined in SPSS (version 20). Because there were multiple tapes that were coded for one observation, these were aggregated for each student by summing the seconds for each activity within a type of instruction (e.g., comprehension, print vocabulary) providing a total amount of instruction for each student.

The amounts of each type of instruction were then combined based on our theory of reading instruction (Connor, Morrison, & Katch, 2004; Connor, Morrison, & Petrella, 2004). Meaning-focused instruction was comprised of all instructional activities that might be expected to *explicitly* support students' language and comprehension skill gains including text reading, writing, oral language, listening and reading comprehension, vocabulary, and other meaning-focused types of instruction (see Table 1). There were also activities that might be expected to support language and comprehension more *implicitly* through instruction in how to decode unfamiliar words and building automaticity, which might be considered more code-focused types of instruction. Descriptive statistics for each type of instruction are provided in Table 3. Of note, most of the time was spent in teacher/child-managed whole class meaning-focused instruction (12.6 minutes per day) followed by teacher/child-managed small-group meaning-focused (9.32 minutes per day). Children spent about 11 minutes per day working with peers or individually on meaning-focused activities. Less time was spent in code-focused activities, about 6 minutes per day compared to a total of about 19 minutes in meaning-focused activities.

### **How does combining measures of the duration of different types of instruction and content for individual students with a more global measure of the CLE synergistically affect students' reading comprehension and vocabulary outcomes?**

As noted in the methods, we ran three different models with the spring VocComp score as the outcome and controlling for fall VocComp scores. The ICC for the unconditional model, which is the between-classroom variance explained, was .24, indicating that approximately 24% of the differences in scores among students were related to the classroom to which they were assigned.

### **Main Effects of Classroom Learning Environment and Amounts, Types, and Content of Instruction on Students' Vocabulary and Reading Comprehension Outcomes**

We then examined the effect of the CLE score on students' spring VocComp score. There was a trend that CLE predicted students' spring outcomes ( $p = .064$ ). A one-point increase in CLE quality score was associated with about a .02 increase in the VocComp spring score ( $d = .044$ , which is negligible). The model explained about 81% of the variability in children's scores compared to the unconditional model.

Next, we removed the CLE quality score from the model and added all eight of the student amount/type/content of instruction time. Results revealed that none of the instruction duration variables predicted students' scores. The model explained approximately 80% of the variance in student scores compared to the unconditional model.

### **Classroom Learning Environment Quality X Duration/Type/Content Interactions**

We then added the CLE quality score to the model and tested all of the CLE quality X time/type/content of instruction interactions (see Table 4). We trimmed three variables (i.e., teacher/child-managed code-focused small-group and whole-class instruction, and child-managed small-group code-focused instruction) and the interactions that did not significantly predict student outcomes. We also tested for student fall VocComp X instruction interaction effects. None significantly predicted spring VocComp outcomes (e.g., fall VocComp X CLE quality interaction effect coefficient =  $-.008$ ,  $p = .400$ ) and so were trimmed from the model. This model explained about 81.3% of the variability in students' scores.

Our models revealed a number of global CLE quality X individual student-level instruction amount/content interactions that significantly predicted students' outcomes (see Table 4 and Figure 2) regardless of students' fall VocComp z-score. Specifically, for teacher/child-managed small-group *and* whole-class meaning-focused instruction, when provided by teachers who were judged to be providing a higher quality CLE, the effect for students who spent more time in meaning-focused instruction was much greater, minute for minute, than the same amount of instruction provided by teachers who were judged to be providing a CLE of lesser quality. The effect was substantial. For example, a student who received 18 minutes of teacher/child-managed small-group meaning-focused instruction and whose teachers received a CLE score of 17 (75<sup>th</sup> percentile of the sample) would achieve scores that were .43 z-score points ( $M = 0$ ,  $SD = 1$ ) higher than a student who received the same amount of instruction but whose teachers received a rating of 13 (25<sup>th</sup> percentile of the

sample), with an effect size ( $d$ ) of .43. The difference for teacher/child-managed whole-class meaning-focused instruction was smaller, .19 z-score points ( $d = .19$ ) or about half of the small-group effect size.

## Discussion

Overall, the teachers in this study provided generally high quality classroom learning environments, but there was substantial variability with two teachers providing CLEs that were judged to be very low – a 6 or worse (out of 18); they received no more than a 1, 2 or 3 on all three scales. At the same time, six teachers had almost perfect scores of 15, having received 5s and 6s on all three scales. There was similar variability in the amount of time third graders spent in teacher/child-managed meaning-focused instruction both within and between classrooms. Neither CLE quality nor the amount-content-type of instruction (ISI/Pathways) individual students received independently predicted students' vocabulary and comprehension gains. Instead, the two systems synergistically captured the complexity of classroom instruction at the individual student level and the more global classroom level. Teachers judged to provide a high quality CLE but whose students received very little teacher/child-managed meaning-focused instruction (e.g., less than 1 minute, see Figure 2), were no more effective than those judged to provide a low quality CLE. As students spent more time in teacher/child-managed meaning-focused small-group and whole-class instruction, differences in low versus high quality CLE effects became larger. Moreover, because students who shared a classroom experienced different amounts of small-group teacher/child-managed meaning focused instruction, using both systems helped to explain within-classroom differences in students' outcomes, which classroom-level systems obscure.

Students showed the greatest gains in vocabulary and comprehension when their teachers provided a high quality CLE *and* they spent greater amounts of time (e.g., 25-35 min, see Figure 2) in teacher/child-managed meaning-focused instruction, particularly when this instruction was provided in small groups. Teacher/child-managed small-group meaning-focused instruction was more than twice as effective as whole-class instruction. When we tested students' fall score X instruction interactions, none significantly predicted spring outcomes. This indicates that results were similar for students regardless of fall vocabulary and comprehension scores.

Another consideration is that for this sample of students, teachers, and classrooms, student characteristic X instruction time-type-content effects on students' literacy gains were documented (Connor, Morrison, Fishman, et al., 2011). Specifically, using the ISI/Pathways system, the smaller the difference between the observed amount of a particular type-content of instruction and the recommended amount for a particular student based on his or her assessed vocabulary and literacy skills (i.e., distance from recommendation), the greater were his or her reading comprehension skill gains.

These findings support a complex systems model of how individual children learn in the context of classroom literacy instruction and, in combination with other studies (Connor, Morrison, Fishman, et al., 2011), extend our understanding of classroom instruction and learning environments as dynamic systems with interacting effects (Yoshikawa & Hsueh,

2001). This indicates that we will be more likely to identify key aspects of complex teaching and classroom-learning environments by using multiple frameworks and considering potential interactions among sources of influence at both the more micro student level as well as the more global classroom level. Thus a student in a classroom-learning environment that is generally judged to be of high quality may not show achievement gains because the student is not receiving appropriate amounts of the particular types and content of instruction that would support his or her achievement. Indeed, students in this study who were judged to be in high quality CLE were no more likely to participate in substantial amounts of teacher/child-managed meaning-focused instruction than were students in low quality CLE. As one reviewer noted, “These are the well-organized, “nice” classrooms that [some] principals and parents love, even though the environment while pleasant is not particularly effective.”

Such interaction effects may help to explain equivocal findings when only one system of observation is utilized and individual student differences are not considered. For example, in the Kane et al. study, (2012), one reason that the assessments of CLE might not have predicted language arts outcomes was because, even in classrooms with teachers judged to be of high quality, students might not have received adequate amounts of specific types of content instruction tailored to their learning needs. None of the observation systems used in the Kane study considered time in content and type of literacy instruction at the student level or student characteristic X instruction interaction effects. At the same time, systems that consider a single time point in certain types-content of literacy instruction and that do not consider quality or child X instruction interactions may not be generally highly predictive either because more time spent in low quality unaligned instruction is unlikely to positively predict student achievement.

On average, students spent about only five minutes per day engaged in oral language and vocabulary instruction, which by any standard, is a minimal amount of time. This implies that there was very little time for explicit instruction or other types of vocabulary instruction (e.g., discussing or clarifying words) to take place. These findings are in line with previous research, which has demonstrated that in many elementary school classrooms, there is a very limited focus on vocabulary instruction (Biemiller, 2001; Durkin, 1979; Scott & Nagy, 1997). Robust vocabulary instruction (Beck et al., 2002) requires teachers to allot sufficient amounts of time that can be used for explaining, providing examples, and elaborating on vocabulary knowledge in order to promote greater understanding. Five minutes per day of vocabulary instruction seems hardly adequate.

A closer look at the types of comprehension instruction provided (see Appendix C), revealed that teachers used over 20 different kinds of teacher/child-managed whole class comprehension instructional activities (see Table C.1) and fewer (about 18) kinds of activities in small groups (Table C.2) to build comprehension. The most salient types observed involving the whole class were *questioning* (3.1 min), *highlighting/identifying* (about 1 min) and *schema building* (about 45 sec) (see Appendix B). In teacher/child-managed small-group instruction, again, the most salient activity observed was *questioning* (1.6 min) followed by *compare and contrast* (13 seconds) and the use of *graphic organizers* (8 sec). In both whole class and small group contexts, amounts varied widely (see Table C.

1). For example, *questioning* ranged from 0 to 22 min and use of graphic organizers ranged from 0 to 16 min.

In general, the comprehension instruction observed was aligned with the findings of the National Reading Panel report (NICHD, 2000) with a focus on strategies (compare and contrast; highlighting; graphic organizers) but little support for more complex understanding of text, which is now required by the Common Core State Standards. We consider questioning among the most basic tools the teacher might use to build reading for understanding (Cazden, 1988). In contrast, very little time was spent on higher-level inferencing in either whole class (about 15 seconds summing across types for whole class) or small group (about 12 seconds). Research shows that inferencing is associated with successful comprehension (Cain et al., 2004; Cromley & Azevedo, 2007) and is a core principal of the Common Core State Standards (Common Core State Standards Initiative, 2010).

One aspect of the study that deserves further investigation is whether dimensions of the CLE might be better predictors when coded at the level of the individual student. It is conceivable that Student A and Student C might be participating in the same amount of time in appropriate learning opportunities but that the teacher is interacting with Student C in ways that are more responsive than with Student A. Hence, one might hypothesize that Student C will demonstrate stronger achievement than will Student A. Measuring CLE and instruction at the level of the individual student is time consuming but may be worth the effort when trying to understand learning and development in classrooms. Another consideration is that dimensions of the CLE quality measure captured the social and emotional climate of the classroom as well as the learning environment. It may be that, particularly for children from low-income families, this more nurturing aspect of the classroom environment is providing a safe-haven that facilitates learning albeit indirectly. There is evidence of this effect in preschool, kindergarten, and first grade classrooms (Hamre & Pianta, 2005; Rimm-Kaufman et al., 2005) as well as for older children (Reyes, Brackett, Rivers, White, & Salovey, 2012).

### Limitations

There are limitations to this study that should be considered when interpreting these results. First, all participating teachers received professional development. This study was conducted in the context of a randomized controlled trial where there was a significant effect of the individualized reading compared to the vocabulary intervention (Connor, Morrison, Fishman, et al., 2011). Although there were no differences between conditions on the quality of the CLE, students who were in classrooms where teachers learned to individualize instruction were more likely to participate in teacher/child-managed small-group meaning-focused instruction that was individualized to their learning needs (Connor, Morrison, Fishman, et al., 2011). The randomized control trial may have influenced the results presented here, and it is possible that our results may not generalize to classrooms in which teachers do not receive professional development. It might also explain why there were no significant student fall VocComp score X instruction interaction effects on spring outcomes. In another sample, such interaction effects might influence achievement. Additionally, the three observations were scheduled at the teachers' convenience, so we most likely observed

higher quality instruction than might have been observed otherwise. Use of more frequent observations would have improved reliability (Rowan & Correnti, 2009) but were not feasible within the funding and time constraints of this study.

### Practical Implications

This study provides insight into the amounts, content, and types of instruction in which individual students participate and qualitative aspects of the classroom-learning environment that appear to be more effective for improving students' literacy and language outcomes. Teacher/child-managed meaning-focused but not teacher/child-managed code-focused instruction predicted third-graders' vocabulary and comprehension achievement. This might not be unexpected inasmuch as explicit instruction in the target outcome tends to be a better predictor than more implicit or indirect instruction (Connor, Morrison, & Katch, 2004; Connor, Morrison, & Petrella, 2004; Connor, Morrison, & Slominski, 2006), and the outcomes were specifically meaning-focused. Small-group instruction was twice as effective as whole-class, perhaps because the teacher was better able to tailor instruction to meet the learning needs of students when interacting with smaller numbers of students. Additionally, he or she was likely to be more responsive, which the extant literature has established as important for student learning (Mashburn et al., 2008). This responsiveness was a key dimension of the CLE rubric. Perhaps, the most important finding was that type, amount, and content of instruction individual students received *and* the quality of the classroom learning environment matter – Students should learn best when provided enough time in explicit instruction from the teacher that is interactive, responsive, organized, and focused on providing targeted language and literacy content in ways that facilitate language and vocabulary learning.

Classroom observation systems represent an important move toward policy that promises to make a true difference in what is defined as high quality and effective teaching, what it looks like in the classroom, and how these practices can be more widely disseminated so that all students, including the most vulnerable, can experience effective instruction and academic gains. One challenge will be designing systems that can be used validly and reliably by school professionals (Crawford et al., in press; Reddy et al., in press) who have varying levels of expertise and knowledge about literacy development. By better understanding the affordances of teaching and the classroom-learning environment that contribute to individual student's language and literacy development, we can design more effective instructional regimens, identify effective standards of practice, discover better ways to measure effective teaching, and develop targeted professional development for teachers and educational leaders that will ensure that all children have the opportunity to learn.

### Acknowledgments

We thank Dr. Elizabeth Crowe, Dr. Stephanie Day, Dr. Jennifer Dombek, and the ISI Project team members for their hard work providing professional development, collecting data, and coding video. We thank the children, parents, teachers, and school administrators without whom this research would not have been possible. This study was funded by grants R305H04013 and R305B070074, "Child by Instruction Interactions: Effects of Individualizing Instruction" from the US. Department of Education, Institute of Education Sciences and by grants R01HD48539, R21HD062834 and P50 HD052120 from the Eunice Kennedy Shriver National Institute of Child



Health and Human Development. The opinions expressed are ours and do not represent views of the funding agencies.

## Appendix A

## The ISI Classroom Learning Environment Scale (excerpts – the full scale is available upon request)

Classroom Orienting, Organization and Planning	Support for Language/Vocabulary	Warmth and Responsiveness/Control/Discipline
<p><b>Rating 1 Indicators</b></p> <p>No evidence of classroom organization. Teacher frequently does not have materials ready or enough materials for all children. Classroom is frequently chaotic and very little time is spent on meaningful instruction. No observable system is in place to facilitate students' transition from one station or location to another.</p>	<p><b>Rating 1 Indicators</b></p> <p>Teacher does not introduce any new words, does not provide explicit or systematic instruction in vocabulary and does not provide opportunities for students to engage in oral language. Students are not provided opportunities to practice key vocabulary. Teacher does not monitor students' vocabulary and comprehension.</p>	<p><b>Rating 1 Indicators</b></p> <p>No evidence that the teacher redirects in respectful ways nor is there evidence that the teacher emphasizes student change in behavior through praise. There is no evidence of the teacher communicating what students did correctly or how they can improve. There is no evidence of students treating each other with respect. Whenever discipline is imposed, it is ineffective.</p>
<p><b>Rating 3 Indicators</b></p> <p>Transitions are of reasonable length but not consistently efficient (not all children). There is an observable, but not always efficient or working system (e.g., center chart, daily schedule) in place for organizing students into groups. The teacher may use a daily lesson plan (e.g., group activity planner print-out).</p>	<p><b>Rating 3 Indicators</b></p> <p>Teacher introduces too many new Tier 1 words per story/text and not enough Tier 2 words. Provides explicit or systematic vocabulary instruction (not both). Occasionally extends meanings. Occasionally provides opportunities for students to engage in oral language and practice key vocabulary. Teacher monitors students' vocabulary and comprehension, but rarely provides feedback.</p>	<p><b>Rating 3 Indicators</b></p> <p>Teacher inconsistently redirects in respectful ways and inconsistently emphasizes student change in behavior through praise. Teacher talk is inconsistently encouraging and respectful and inconsistently connects students' personal experiences to lesson content. Inconsistently communicates clearly what students did correctly or how they can improve. Students inconsistently treat each other with respect.</p>
<p><b>Rating 6 Indicators</b></p> <p>The classroom is well organized and instruction is well organized. Classroom routine is evident. Transitions are efficient.</p>	<p><b>Rating 6 Indicators</b></p> <p>Word knowledge is an ongoing part of the instructional day. Teacher's selection of words demonstrates knowledge of the words' utility and relation to previously known words and relevance for text being taught. Students are encouraged to make connections between words/meaning they are already familiar with and new words/meanings. Much of the instruction is provided in small groups.</p>	<p><b>Rating 6 Indicators</b></p> <p>Teacher is the authority figure in the class but is never punitive. Classroom consistently offers a positive learning environment with clear expectations for students' behavior as a member of the learning community. Effectively selects and incorporates students' responses, ideas, examples, and experiences into the lesson.</p>

## Appendix B

### Listening and Reading Comprehension Excerpts from the ISI/Pathways Coding Manual (full manual available upon request)

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 or listening of connected text (e.g., silent sustained reading followed by a comprehension worksheet, comprehension strategy instruction using a particular example of connected text, an interactive teacher read aloud during which the teacher models various comprehension strategies).

#### 7.1.10.3 Schema and Concept Building (Modifier)

Listening and Reading Comprehension>Schema Building should be coded for activities which 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 which 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.

#### 7.1.10.6 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 says, "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 vs. 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 which 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 by 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 AR tests which are typically completed on the computer; AR test should also be coded as event code > Assessment. This code should also be used a default code for activities where it is not clear whether activity is highlighting, questioning, or summarizing.

## Appendix C

**Table C.1**

Descriptive Statics for Teacher/Child-Managed Comprehension Whole-class Instruction in Seconds

	Min	Max	Mean	SD
Previewing	.00	374.73	24.42	72.80
Schema building	.00	588.29	46.60	112.47
Questioning	.00	1772.7	186.0	294.66
Monitoring	.00	235.49	7.15	29.59
Highlighting/Identifying	.00	985.10	55.91	126.84
Context Cues	.00	327.01	8.04	37.45
Graphic/Semantic Organizers	.00	725.37	12.54	86.33
Prior Knowledge	.00	369.65	19.24	52.31
Retell	.00	192.24	7.75	29.95
Sequencing	.00	168.71	1.03	12.81
Compare/Contrast	.00	380.58	22.77	73.03
Comprehension other	.00	847.11	17.72	70.53
Multicomponent integrated Comprehension Strategies	.00	808.00	19.91	100.80
Predicting	.00	109.56	9.93	24.28
Inferencing Between Texts	.00	39.39	.01	4.46
Inferencing Background Knowledge	.00	180.55	5.38	23.98
Inferencing Within Text	.00	320.21	8.92	34.85
Summarizing Main Idea	.00	129.28	3.16	15.95
Fact vs. Opinion	.00	159.19	2.06	18.05
Cause and Effect	.00	327.01	3.98	31.47
Narrative text	.00	539.77	8.71	62.59
Expository text	.00	118.34	1.15	11.64

**Table C.2**

Descriptive Statistics for Teacher/Child-Managed Comprehension Small-Group Instruction in Seconds

Type of Comprehension Instruction	Mini	Max	Mean	SD
Previewing	.00	266.06	2.10	18.60
Schema building	.00	248.71	6.13	27.88
Questioning	.00	1338.96	95.26	217.63
Monitoring	.00	111.81	.36	6.37
Highlighting Identifying	.00	169.00	7.22	27.02
Context Cues	.00	178.70	2.49	17.17
Graphic/Semantic Organizers	.00	989.42	8.03	81.57
Prior Knowledge	.00	103.43	2.09	9.42
Retell	.00	33.63	.10	1.91
Sequencing	.00	36.39	.28	2.90
Compare/Contrast	.00	794.63	13.27	89.27
Multicomponent Integrated Comprehension Strategies	.00	420.09	6.30	41.95
Predicting	.00	569.08	9.03	65.10
Inferencing Background Knowledge	.00	333.89	2.74	26.83
Summarizing/Main Idea	.00	196.00	2.47	18.84
Fact vs. Opinion	.00	629.78	5.67	51.18
Cause and Effect	.00	176.17	1.08	13.43
Narrative text	.00	14.77	.04	.84
Workbook worksheet	.00	1016.78	9.57	77.19
Narrative Text Language Arts	.00	919.15	48.41	142.05
Expository Text Language Arts	.00	415.08	5.82	46.34
Expository Text Science	.00	560.82	11.59	65.63
Workbook Worksheet Language	.00	989.11	49.18	158.23
Workbook Worksheet SocialStudies	.00	174.10	1.13	14.00
Blackboard Language Arts	.00	54.20	.17	3.08
Expository Text Social Studies	.00	255.99	.94	14.71
Workbook Worksheet Science	.00	575.05	10.32	60.75
Journal	.00	137.78	.44	7.85
Summarizing	.00	196.00	3.28	19.80
Inferencing Within Text	.00	294.22	9.48	35.59

## References

- Beck, IL.; McKeown, Margaret G.; Kucan, Linda. solving problems in the teaching of literacy. Guilford Publications; New York: 2002. Bringing words to life: Robust vocabulary instruction..
- Beck IL, Perfetti CA, McKeown MG. Effects of long term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*. 1982; 74(4):506–521.
- Biemiller, Andrew; Boote, Catherine. An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology*. 2006; 98(1):44–62.

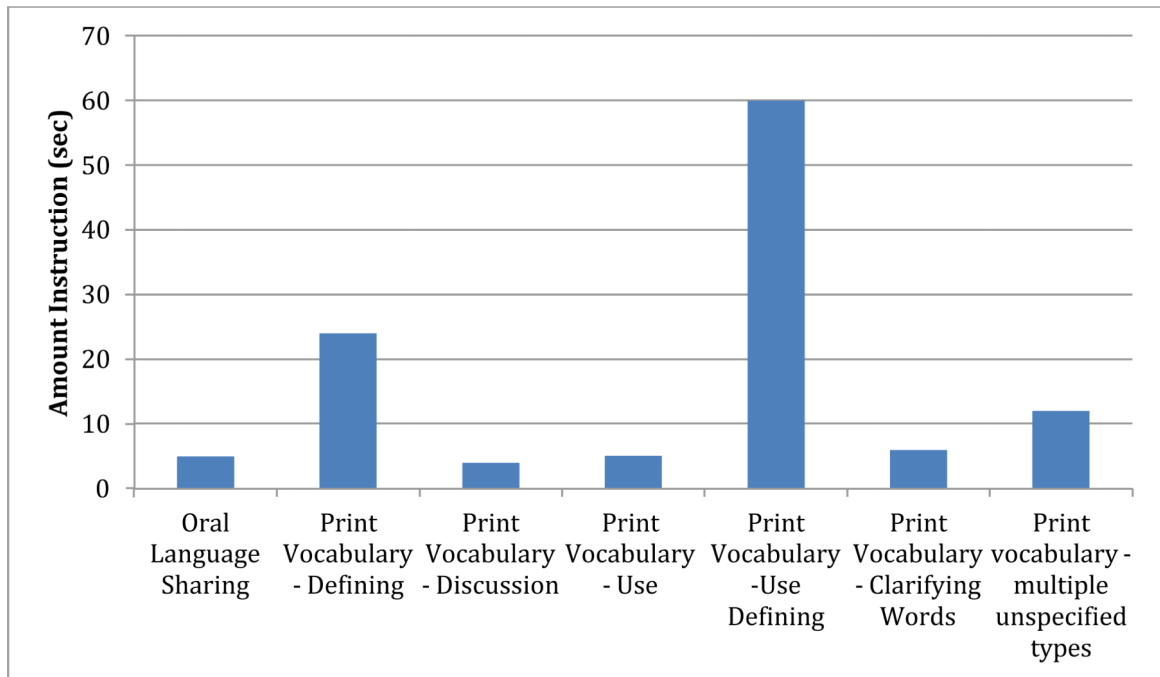
- Bogdan, Robert C.; Biklen, Sari Knopp. *Qualitative research in education An introduction to theory and method*. 3rd ed.. Allyn & Bacon; Needham Heights, MA: 1998.
- Borko, H.; Niles, J. Descriptions of teacher planning: Ideas for teachers and research.. In: Richardson-Koehler, V., editor. *Educators' handbook: A research perspective*. Longman; New York: 1987. p. 167-187.
- Bronfenbrenner, U.; Morris, PA. The bioecological model of human development.. In: Lerner, RM.; Damon, W., editors. *Handbook of child psychology: Theoretical models of human development*. 6th ed.. Vol. 1. John Wiley & Sons; Hoboken, NJ: 2006. p. 793-828.
- Brophy, Jere E. Teacher behavior and its effects. *Journal of Educational Psychology*. 1979; 71(6):733–750.
- Cain, Kate; Oakhill, Jane; Lemmon, Kate. Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*. 2004; 96(4):671–681.
- Cameron, Claire E.; Connor, Carol McDonald; Morrison, Frederick J. Effects of variation in teacher organization on classroom functioning. *Journal of School Psychology*. 2005; 43(1):61–85.
- Cameron, Claire E.; Connor, Carol McDonald; Morrison, Frederick J.; Jewkes, AM. Effects of Classroom Organization on Letter-Word Reading in First Grade. *Journal of School Psychology*. 2008; 6(2):173–192. [PubMed: 19083356]
- Carlisle JF. Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading & Writing*. 2000; 12(3-4):169–190.
- Cassidy J, Cassidy D. What's hot, what's not for 2006. *Reading Today*. 2005/2006; 23:1.
- Cazden, C. *Classroom discourse*. Heineman; Portsmouth, NH: 1988.
- Chall, JS. *Learning to read: The great debate*. McGraw-Hill Book Co.; New York: 1967.
- Common Core State Standards Initiative. [March, 2013] *Common Core State Standards for Mathematics*. 2010. from [http://www.corestandards.org/assets/CCSSI\\_Math\\_Standards.pdf](http://www.corestandards.org/assets/CCSSI_Math_Standards.pdf)
- Connor, Carol McDonald. Commentary on Two Classroom Observation Systems: Moving Toward a Shared Understanding of Effective Teaching. *School Psychology Quarterly*. 2013a; 28(4):342–346. doi: 10.1037/spq0000045. [PubMed: 24341927]
- Connor, Carol McDonald. Intervening to support reading comprehension development with diverse learners.. In: Miller, B.; Cutting, LE., editors. *Unraveling the Behavioral, Neurobiological and Genetic Components of Reading Comprehension: The Dyslexia Foundation and NICHD*. Brookes; Baltimore: 2013b. p. 222-232.
- Connor, Carol McDonald; Morrison, Frederick J. Knowledge Acquisition in the Classroom: Literacy and Content Area Knowledge. In: Pinkham, AM.; Kaefer, T.; Neuman, SB., editors. *Knowledge Development in Early Childhood: How Young Children Build Knowledge and Why It Matters*. Guilford Press; New York: 2012. p. 220-241.in press
- Connor, Carol McDonald; Morrison, Frederick J.; Fishman, Barry; Crowe, Elizabeth C.; Al Otaiba, Stephanie; Schatschneider, Christopher. A longitudinal cluster-randomized control study on the accumulating effects of individualized literacy instruction on students' reading from 1st through 3rd grade. *Psychological Science*. 2013; 24(8):1408–1419. doi: 10.1177/0956797612472204. [PubMed: 23785038]
- Connor, Carol McDonald; Morrison, Frederick J.; Fishman, Barry; Giuliani, Sarah; Luck, Melissa; Underwood, Phyllis; Schatschneider, Christopher. Classroom instruction, child X instruction interactions and the impact of differentiating student instruction on third graders' reading comprehension. *Reading Research Quarterly*. 2011; 46(3):189–221.
- Connor, Carol McDonald; Morrison, Frederick J.; Fishman, Barry; Ponitz, Claire Cameron; Glasney, Stephanie; Underwood, Phyllis; Schatschneider, Christopher. The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher*. 2009; 38(2):85–99.
- Connor, Carol McDonald; Morrison, Frederick J.; Katch, E. Leslie Beyond the reading wars: The effect of classroom instruction by child interactions on early reading. *Scientific Studies of Reading*. 2004; 8(4):305–336.

- Connor, Carol McDonald; Morrison, Frederick J.; Petrella, Jocelyn N. Effective reading comprehension instruction: Examining child by instruction interactions. *Journal of Educational Psychology*. 2004; 96(4):682–698.
- Connor, Carol McDonald; Morrison, Frederick J.; Schatschneider, Christopher; Toste, Jessica; Lundblom, Erin G.; Crowe, Elizabeth; Fishman, Barry. Effective classroom instruction: Implications of child characteristic by instruction interactions on first graders' word reading achievement. *Journal for Research on Educational Effectiveness*. 2011; 4(3):173–207.
- Connor, Carol McDonald; Morrison, Frederick J.; Slominski, Lisa. Preschool instruction and children's literacy skill growth. *Journal of Educational Psychology*. 2006; 98(4):665–689.
- Connor, Carol McDonald; Piasta, Shayne B.; Fishman, Barry; Glasney, Stephanie; Schatschneider, Christopher; Crowe, Elizabeth; Morrison, Frederick J. Individualizing student instruction precisely: Effects of child by instruction interactions on first graders' literacy development. *Child Development*. 2009; 80(1):77–100. [PubMed: 19236394]
- Connor, Carol McDonald; Son, Seung-Hee; Hindman, Annemarie H.; Morrison, Frederick J. Teacher qualifications, classroom practices, family characteristics, and preschool experience: Complex effects on first graders' vocabulary and early reading outcomes. *Journal of School Psychology*. 2005; 43:343–375.
- Crawford, April D.; Zucker, Tricia A.; Williams, Jeffrey M.; Bhavsar, Vibhuti; Landry, Susan H. Initial validation of the pre-kindergarten classroom observation tool and goal setting system for data-based coaching. *School Psychology Quarterly*. in press.
- Cromley, Jennifer G.; Azevedo, Roger. Testing and refining the direct and inferential mediation (DIME) model of reading comprehension. *Journal of Educational Psychology*. 2007; 99(2):311–325.
- Danielson, C. *Enhancing Professional Practice: A framework for teaching*. 2nd ed.. ASCD; Alexandria, VA: 2007.
- Goldhaber, Dan; Anthony, Emily. *Teacher Quality and Student Achievement*. Urban Diversity Series (pp. Report: UDS-115; 153). Department of Education, Washington, DC.; New York: 2003.
- Gottardo, Alexandra; Stanovich, Keith E.; Siegel, Linda S. The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third grade children. *Journal of Experimental Child Psychology*. 1996; 63:563–582. [PubMed: 8954609]
- Grossman, P.; Loeb, S.; Cohen, J.; Hammerness, K.; Wyckoff, J.; Boyd, D.; Lankford, H. Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores (No. 45). National Center for Analysis of Longitudinal Data in Education Research; Washington, DC: 2010.
- Hamre, Bridget K.; Pianta, Robert C. Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*. 2005; 76(5):949–967. [PubMed: 16149994]
- Hamre, Bridget K.; Pianta, Robert C.; Downer, Jason T.; Mashburn, Andrew J. 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. Apr. 2007
- Justice LM, Petscher Y, Schatschneider C, Mashburn A. Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills. *Child Development*. 2011; 82(6):1768–1777. [PubMed: 22026480]
- Kane, Thomas; Staiger, Douglas O.; McCaffrey, Dan. *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains: Bill and Melinda Gates Foundation*. 2012
- Keenan JM, Betjemann RS, Olson RK. Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*. 2008; 12(3):281–300.
- Kline, Rex B. *Principles and practice of structural equation modeling*. The Guilford Press; New York: 1998.
- Konstantopoulos S, Chung N. The persistence of teacher effects in elementary grades. *American Educational Research Journal*. 2011; 48(2):361–386.

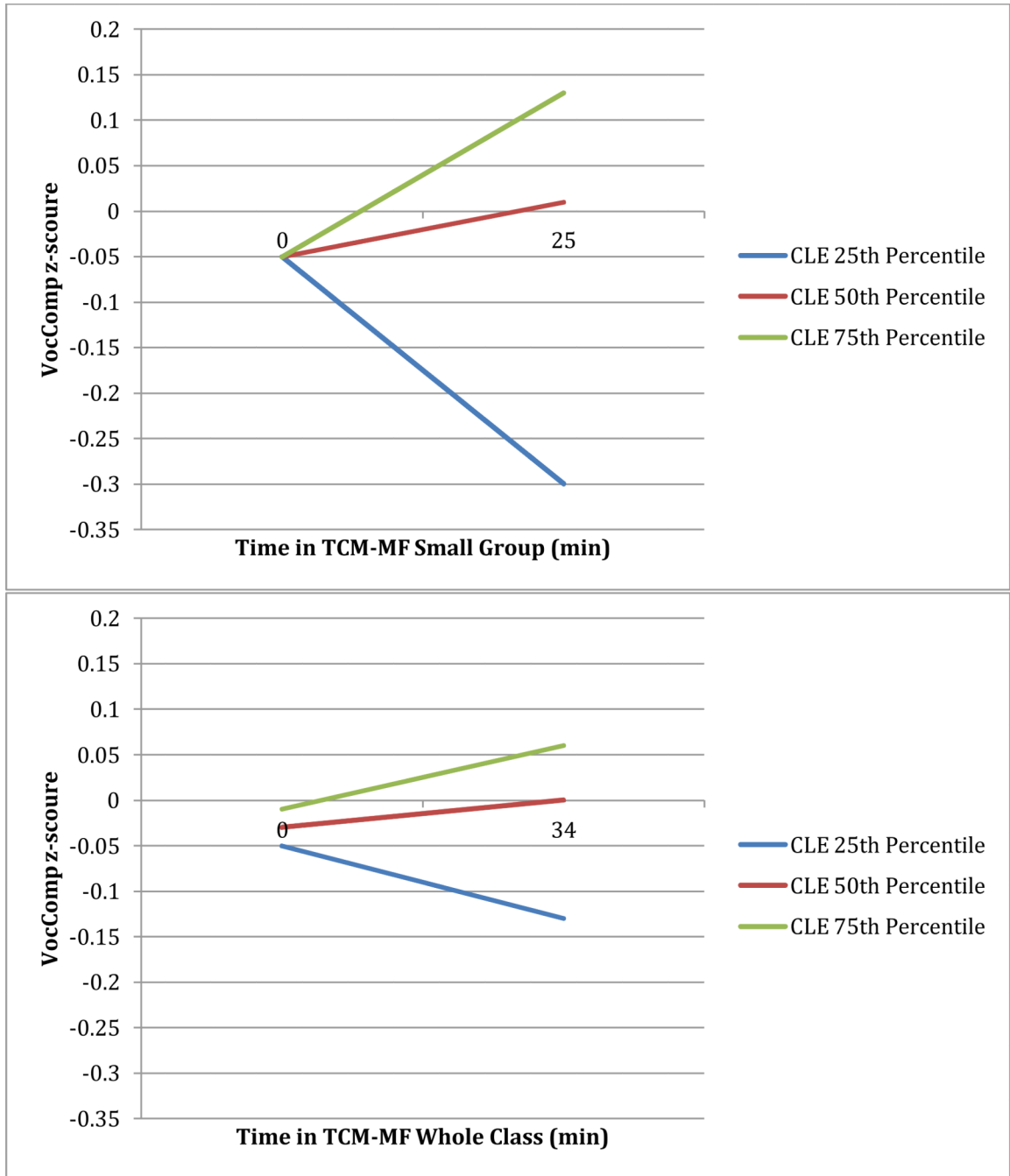
- Landis, J. Richard; Koch, Gary G. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159–174. [PubMed: 843571]
- MacGinitie, WH.; MacGinitie, RK. *Gates-MacGinitie Reading Tests*. 4th ed.. Houghton Mifflin; Iowa City: 2006.
- Mashburn AJ, Pianta RC, Hamre Bridget K, Downer JT, Barbarin O, Bryant D, et al. Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*. 2008; 79(3):732–749. [PubMed: 18489424]
- Mehta PD, Foorman Barbara R, Branum-Martin L, Taylor WP. Literacy as a Unidimensional Multilevel Construct: Validation, Sources of Influence, and Implications in a Longitudinal Study in Grades 1 to 4. *Scientific Studies of Reading*. 2005; 9(2):85–116.
- National Center for Education Statistics, Institute of Education Sciences, US Department of Education. *Fast Facts: English Language Learners*. Institute of Education Sciences, US Department of Education; 2013. from <http://nces.ed.gov/fastfacts/display.asp?id=96> [December]
- NICHD. National Institute of Child Health and Human Development, National Reading Panel report: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development; Washington DC: 2000.
- Pianta, Robert C.; Belsky, Jay; Houts, R.; Morrison, Frederick J.; NICHD-ECCRN. TEACHING: Opportunities to learn in America's elementary classrooms. *Science*. 2007; 315:1795–1796. March 30. [PubMed: 17395814]
- Pianta, Robert C.; La Paro, Karen M.; Payne, K.; Cox, C.; Bradley, RH. The relation of kindergarten classroom environment to teacher, family and school characteristics and child outcomes. *Elementary School Journal*. 2002; 102(3):225–238.
- Piasta, Shayne B.; Connor, Carol McDonald; Fishman, B.; Morrison, Frederick J. Teachers' knowledge of literacy, classroom practices, and student reading growth. *Scientific Studies of Reading*. 2009; 13(3):224–248.
- Ramey, SL.; Ramey, C. Early educational interventions: Principles of effective and sustained benefits from targeted early education programs.. In: D. D. S. B.; Neuman, SB., editors. *Handbook of early literacy research*. Vol. 2. Guilford; New York, NY: 2006. p. 445-459.
- Rasch G. *Winsteps (Version 3.30)* [Statistics]. 2001
- Raudenbush, Stephen W. The Brown Legacy and the O'Connor Challenge: Transforming Schools in the Images of Children's Potential. *Educational Researcher*. 2009; 38(3):169–180. doi: 10.3102/0013189x09334840.
- Raudenbush, Stephen W.; Bryk, Anthony S. *Hierarchical linear models: Applications and data analysis methods*. second ed.. Sage; Thousand Oaks, CA: 2002.
- Raver C, Cybele Blair, C. Li-Grining CP. Extending models of emotion self-regulation to classroom settings: Implications for professional development. 2011 in review.
- Reddy, Linda A.; Fabiano, Gregory; Dudek, Christopher; Hsu, Louis. Predictive validity of the Classroom Strategies Scale–Observer Form on statewide testing scores: an initial investigation. *School Psychology Quarterly*. in press.
- Reis, Sally M.; McCoach, D. Betsy; Little, Catherine A.; Muller, Lisa M.; Kaniskan, R. Burcu The Effects of Differentiated Instruction and Enrichment Pedagogy on Reading Achievement in Five Elementary Schools. *American Educational Research Journal*. 2011; 48(2):462–501. doi: 10.3102/0002831210382891.
- Reyes, Maria R.; Brackett, Marc A.; Rivers, Susan E.; White, Mark; Salovey, Peter. Classroom Emotional Climate, Student Engagement, and Academic Achievement. *Journal of Educational Psychology*. No Pagination Specified. 2012 doi: 10.1037/a0027268.
- Reynolds AJ, Magnuson KA, Ou Suh-Ruu. Preschool-to-third grade programs and practices: A review of research. *Children and Youth Services Review*. 2010; 32:1122–1130.
- Rimm-Kaufman, Sara E.; Paro, Karen M. La; Downer, Jason T.; Pianta, Robert C. The contribution of classroom setting and quality of instruction to children's behavior in kindergarten classrooms. *Elementary School Journal*. 2005; 105(4):337–345.



- Rowan, Brian; Correnti, Richard. Studying reading instruction with teacher logs: Lessons from the study of instructional improvement. *Educational Researcher*. 2009; 38:120–131.
- Rupley, William H.; Logan, John W.; Nichols, William D. Vocabulary instruction in a balanced reading program. *Reading Teacher*. 1998; 52(4):336–346.
- Snow, CE. Reading for understanding. Santa Monica. RAND Education and the Science and Technology Policy Institute; CA: 2001.
- Snow, CE.; Burns, M. Susan; Griffin, Peg, editors. Preventing reading difficulties in young children. National Academy Press; Washington, DC: 1998.
- Stahl, Steven A. Vocabulary development. United States of America: Brookline Books. 1999
- Taylor, BM.; Pearson, David P., editors. Teaching reading: Effective schools, accomplished teachers. Lawrence Erlbaum; Mahwah, NJ: 2002.
- Taylor BM, Pearson David P. Clark Kathleen, Walpole Sharon. Effective schools and accomplished teachers: lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal*. 2000; 101(2):121–165.
- Therrien WJ. Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education*. 2004; 25(4):252–261.
- Tuyay, Sabrina; Jennings, Louise; Dixon, Carol. Classroom discourse and opportunities to learn: An ethnographic study of knowledge construction in a bilingual third-grade classroom. *Discourse Processes*. 1995; 19:75–110.
- UTOP. 2009. 2013, from [http://uteach.utexas.edu/ResearchMethods/UTOP\\_2009.doc](http://uteach.utexas.edu/ResearchMethods/UTOP_2009.doc)
- Vellutino FR, Tunmer WE, Jaccard J, Chen R, Scanlon DM. Components of reading ability: Multiariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*. 2004; 11(1):3–32. doi: 10.1207/s1532799xssr1101\_2.
- Wharton-McDonald, Ruth; Pressley, Michael; Hampston, Jennifer Mistretta. Literacy instruction in nine first-grade classrooms: Teacher characteristics and student achievement. *Elementary School Journal*. 1998; 99(2):101–128.
- Whitehurst GJ, Falco FL, Lonigan CJ, Fischel JE, DeBaryshe BD, Valdez-Menchaca MC, Caulfield M. Accelerating language development through picture book reading. *Developmental Psychology*. 1988; 24(4):552–559.
- Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson-III Tests of Achievement. Riverside; Itasca, IL: 2001.
- Yoshikawa, Hirokazu; Hsueh, JoAnn. Child development and public policy: Toward a dynamic systems perspective. *Child Development*. 2001; 72(6):1887–1903. [PubMed: 11768151]



**Figure 1.**  
Amounts (sec/day) of TCM small group vocabulary instruction by type.



**Figure 2.** Student-level amounts, content and type of instruction received X Quality of the classroom-learning environment (CLE) Interactions on Vocabulary/Comprehension (VocComp) z-scores. Modeled x-axis 5<sup>th</sup> to 95<sup>th</sup> percentile of time spent in Teacher/child-managed-meaning focused (TCM-MF) Small Group (top) and Whole Class (bottom) instruction, as a function of CLE score modeled at the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles; other variables centered at their mean.

**Table 1**

Dimensions of Instruction (Context, Grouping, and Management) and Content Areas Associated With Code- and Meaning-Focused Types of Instruction

	<b>Code-focused</b>	<b>Meaning-focused</b>
Teacher/child-managed, Whole class	The teacher is teaching the class how to decode multi-syllabic words by writing them on the white board and then demonstrating various strategies, such as looking for prefixes and suffixes.*	The teacher is reading <i>A Single Shard</i> to the class. She stops every so often to ask the students questions.
Teacher/child-managed, Small group	The teacher is working with a small group of children on spelling (i.e., encoding) strategies.	The teacher and a small group of students are discussing <i>Mr. Poppers' Penguins</i> and how it is similar and different from <i>Charlotte's Web</i> .
Child/peer-managed, Small group and individual self-managed	Students are working together in pairs to complete a worksheet on dividing multi-syllabic words into syllables.	Students are working individually to revise an argumentative essay using feedback from their peers.
Content areas (in the coding system)	Phonological Awareness Morphological Awareness* Word Identification/Decoding Word Identification/Encoding Grapheme/Phoneme correspondence Fluency*	Print and Text Concepts Oral Language Print Vocabulary Listening and Reading Comprehension Text Reading Writing

\* *Note.* It can be argued that morphological awareness (Carlisle, 2000) and fluency (Therrien, 2004) might also be considered meaning-focused activities. In our theory of literacy instruction, code-focused activities represent the more automatic processes whereas meaning-focused activities require the integration of the more automatic processes with active construction of meaning of connected text. Hence, code-focused activities are more likely to directly affect aspects of reading related to skill whereas meaning-focused activities are more likely to directly contribute to aspects of comprehension and reading for understanding.

Correlations, Means, Standard Deviations, Standard Scores, Percentile Rank, and Factor Loadings Of the Reading Comprehension and Vocabulary Measures Administered During the Fall and Spring Of the Academic Year

Table 2

	Fall WJ-PC	Fall WJ-PV	Fall GM-C	Fall GM-V	Spring WJ-PC	Spring WJ-PV	Spring GM-C	Spring GM-V
Fall WJ-PC	1							
Fall WJ-PV	.579*	1						
Fall GM-C	.627*	.510*	1					
Fall GM-V	.705*	.669*	.749*	1				
Spring WJ-PC	.697*	.517*	.602*	.663*	1			
Spring WJ-PV	.503*	.821*	.451*	.633*	.523*	1		
Spring GM-C	.622*	.534*	.780*	.740*	.624*	.504*	1	
Spring GM-V	.701*	.674*	.707*	.847*	.670*	.610*	.760*	1
Mean	94.75 <sup>a</sup>	98.92 <sup>a</sup>	51.81 <sup>b</sup>	55.04 <sup>b</sup>	95.96 <sup>a</sup>	99.21 <sup>a</sup>	52.07 <sup>b</sup>	59.84 <sup>b</sup>
SD	10.31 <sup>a</sup>	10.82 <sup>a</sup>	28.04 <sup>b</sup>	27.26 <sup>b</sup>	10.43 <sup>a</sup>	10.79 <sup>a</sup>	29.79 <sup>b</sup>	25.52 <sup>b</sup>
Maximum	124 <sup>a</sup>	127 <sup>a</sup>	99 <sup>b</sup>	99 <sup>b</sup>	121 <sup>a</sup>	139 <sup>a</sup>	99 <sup>b</sup>	99 <sup>b</sup>
Minimum	48 <sup>a</sup>	68 <sup>a</sup>	1 <sup>b</sup>	1 <sup>b</sup>	54 <sup>a</sup>	69 <sup>a</sup>	1 <sup>b</sup>	2 <sup>b</sup>
Factor Loadings	.85	.798	.845	.917	.835	.861	.768	.908

Note: WJ = Woodcock Johnson; PC = passage comprehension; PV = print vocabulary; GM = Gates-MacGinitie; C = comprehension; V = vocabulary; SD = standard deviation.

\*\* Correlation is significant at  $p < .001$ .

<sup>a</sup> Standard scores

<sup>b</sup> Percentile rank.

**Table 3**

Descriptive Statistics for Amount (min) of Student Instruction Variables and Cross Level Associations for Amount and Classroom Learning Environment (CLE) ratings (1 low to 6 high quality)

	<b>N</b>	<b>Mean</b>	<b>SD</b>	<b>Minimum</b>	<b>Maximum</b>
TCM-SG-CF	315	0.82	1.68	0.00	15.05
TCM-SG-MF	315	9.32	9.56	0.00	42.11
CM-SG-MF	315	4.84	6.13	0.00	35.84
CM-SG-CF	315	0.54	1.31	0.00	6.86
CM-WC-CF	315	0.81	1.91	0.00	17.78
CM-WC-MF	315	6.31	6.98	0.00	29.30
TCM-WC-CF	315	2.80	5.18	0.00	31.05
TCM-WC-MF	315	12.61	16.15	0.00	67.81

<b>Student-level Outcome</b>	<b>Classroom-level CLE Quality HLM coefficient (SE)</b>
TCM-SG-MF	.261 (1.67)
TCM-WC-MF	-.386 (1.20)
TCM-SM- CF	.060 (.04)
TCM-WC-CF	-.035 (.23)
CM-SM-MF	.117 (.44)

*Note:* HLM Cross-Level Student and Classroom Associations Unstandardized Coefficients

*Note.* TCM = teacher/child-managed; CM = child-managed; SG = small group; WC = whole class; CF = code-focused instruction; MF = meaning-focused instruction.

**Table 4**

HLM Results: Effect of Classroom Learning Environment (CLE), Amounts/Types/Content of Student Instruction (min), and Interaction Effects on Students' Spring Vocabulary and Comprehension (VocComp) z-scores and HLM model

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
Mean VocComp, $\gamma_{00}$	-0.034	0.032	-1.056	25	0.301
CLE, $\gamma_{01}$	0.027	0.008	3.254	25	0.003
For TCM-SG MF slope, $\beta_1$					
TCM-SG-MF effect, $\gamma_{10}$	-0.003	0.003	-0.751	331	0.453
CLE interaction effect, $\gamma_{11}$	0.005	0.001	3.509	331	<0.001
For CM-SG-MF slope, $\beta_2$					
CM-SG-MF effect, $\gamma_{20}$	0.005	0.005	0.971	331	0.332
CLE interaction effect, $\gamma_{21}$	-0.003	0.001	-1.873	331	0.062
For CM-WC-CF slope, $\beta_3$					
CM-WC-CF effect, $\gamma_{30}$	0.013	0.012	1.102	331	0.271
CLE interaction effect, $\gamma_{31}$	-0.004	0.002	-1.897	331	0.059
For CM-WC-MF slope, $\beta_4$					
CM-WC-MF effect, $\gamma_{40}$	0.001	0.004	0.314	331	0.754
CLE interaction effect, $\gamma_{41}$	0.002	0.001	1.809	331	0.071
For TCM-WC-MF slope, $\beta_5$					
TCM-WC-MF effect, $\gamma_{50}$	0.0001	0.001	0.156	331	0.876
CLE interaction effect, $\gamma_{51}$	0.001	0.000	2.816	331	0.005
For Fall VocComp slope, $\beta_6$					
Fall VocComp effect, $\gamma_{60}$	0.901	0.027	32.755	331	<0.001

Random Effect	SD	Variance Component	d.f.	$\chi^2$	p-value
Classroom Level, $u_0$	0.14574	0.02124	25	50.16340	0.002
Student Level, $r$	0.41352	0.17100			

Deviance = 464.152946

Note. CM = Child Managed, TCM = Teacher/child-managed, SG = small group; WC = whole class, MF = Meaning-focused; CF = Code-focused instruction; SD = standard deviation.

Model: Spring CompVoc<sub>ij</sub> =  $\gamma_{00} + \gamma_{01} * CLE_j + \gamma_{10} * TCM_{SGMFij} + \gamma_{11} * CLE_j * TCM_{SGMFjOBSij} + \gamma_{20} * CMM_{FOBSij} + \gamma_{21} * CLE_j * CMM_{FOBSij} + \gamma_{30} * CM_{WCCFOBij} + \gamma_{31} * CLE_j * CM_{WCCFij} + \gamma_{40} * CM_{WCSGMFij} + \gamma_{41} * CLE_j * CM_{WCMFij} + \gamma_{50} * TCM_{WCMFij} + \gamma_{51} * CLE_j * TCM_{WCMFij} + \gamma_{60} * Fall\ CompVoc\ Factor\ Score_{ij} + \gamma_{61} * CLE_j * Fall\ CompVoc_{ij} + u_{0j} + r_{ij}$