CHILD CHARACTERISTICS BY SCIENCE INSTRUCTION INTERACTIONS IN SECOND AND THIRD GRADE AND THEIR RELATION TO STUDENTS' CONTENT-AREA KNOWLEDGE, VOCABULARY, AND READING SKILL GAINS

ABSTRACT

The associations among second- and third-grade students' content-area knowledge, vocabulary, and reading gains and the science instruction they received were examined in this exploratory longitudinal study. We also asked whether there were child characteristics × instruction interaction effects on students' content-area literacy. Second graders ($n = 88$) were followed into third grade ($n = 73$). Classrooms were observed all day in the fall, winter, and spring, and amounts and types of science instruction and language arts instruction were recorded. Results revealed that specific types of science instruction were related to second- and third-grade students' gains in content-area literacy skills and that the relation of science instruction to outcomes depended on students' fall content-area knowledge, vocabulary, and reading skills. These results suggest that science instruction may promote students' developing content-area literacy growth and may be more effective when it is implemented taking individual student differences into account.

Carol McDonald Connor
Diana C. Rice
Angela I. Canto
Sherry A. Southerland
Phyllis Underwood
FLORIDA STATE UNIVERSITY

Sibel Kaya
KOCAELI UNIVERSITY

Barry Fishman
Frederick J. Morrison
UNIVERSITY OF MICHIGAN

APPROXIMATELY two-thirds of children fail to read at or above proficient levels by fourth grade (National Assessment of Educational Progress [NAEP], 2007), and this rate is higher for children who live in poverty. Proficient readers have been described as students "who are capable of acquiring new knowledge and understanding new concepts, are capable of applying textual information appropriately, and are capable of being engaged in the reading

 0013-5984/2012/11301-0005 $10.00

process and reflecting on what is being read" (Snow, 2002, p. xiii). Thus, being able to decode words is necessary but not sufficient to define a proficient reader. Whereas research has documented ways to improve students' decoding and word-reading skills (National Institute of Child Health and Human Development [NICHD], 2000), there is less evidence on ways to improve students' literacy when using a "reading to learn" definition of proficient reading (Chall, 1996) and when reading and content-area instruction are intertwined (Snow, 2002). One particularly promising context for developing proficient reading skills is the teaching of science (Guthrie et al., 2004; Morrow, Pressley, Smith, & Smith, 1997). It is possible that beginning to encourage literacy development in the rich context of science instruction might also support beginning readers (Williams et al., 2005). Yet, individual differences in children's response to science instruction have not been considered to date, although such differences are well documented for learning to read (Connor, Piasta, et al., 2009; Torgesen, 2000).

Hence the current exploratory study has two principal aims. First, we examined whether second- and third-grade science instruction might be associated with students' content-area literacy gains, taking the amount of language arts instruction they received into consideration. Second, we investigated whether there might be child characteristics × science instruction (CXI) interactions (also called aptitude × treatment interactions; Cronbach & Snow, 1977). Accumulating evidence shows that the effect of reading instruction depends on children's characteristics and, particularly, their language and literacy skills (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; Connor, Piasta, et al., 2009; Juel & Minden-Cupp, 1998). Thus, children may fail to achieve proficient reading skills only because they do not receive the amounts and types of reading instruction that are appropriate for them. The aspects of language and literacy most likely to be predicted by science instruction—content knowledge, vocabulary, and comprehension—are also important child characteristics with regard to instruction (Connor et al., 2007). Thus, the second goal of this study is to examine whether these CXI interaction effects are evident for science instruction. Such findings would have implications for how we teach science and literacy more effectively for children with widely differing content knowledge and literacy skills.

In this study, we investigate the relation between science instruction and students' content-area literacy gains, as informed by the model provided in Figure 1. Incorporating the three elements outlined in the Rand Report—(*a*) the reader, (*b*) the text, and (*c*) the activity (Snow, 2001, p. 12)—the model is also informed by ecological and dynamical systems theories (Bronfenbrenner & Morris, 2006) as well as theories of comprehension (Van den Broek, 2010). Using this framework, students (the readers) bring certain characteristics, including language, vocabulary, knowledge, motivation, and culture, to the text they read and the activities in which they participate. The text and the science activity are interrelated and represent a learning opportunity (Tuyay, Jennings, & Dixon, 1995). These learning opportunities, in turn, influence students' content-area literacy skill gains. For the purposes of this study, we define content-area literacy as including three constructs: content-area knowledge, vocabulary (which is part of the semantic/lexical system), and reading, represented as word-reading skills in second grade and comprehension in third grade. Note that the learning opportunities, activity, and text do not directly relate to students' literacy outcomes but rather act to influence children's trajectory of literacy development

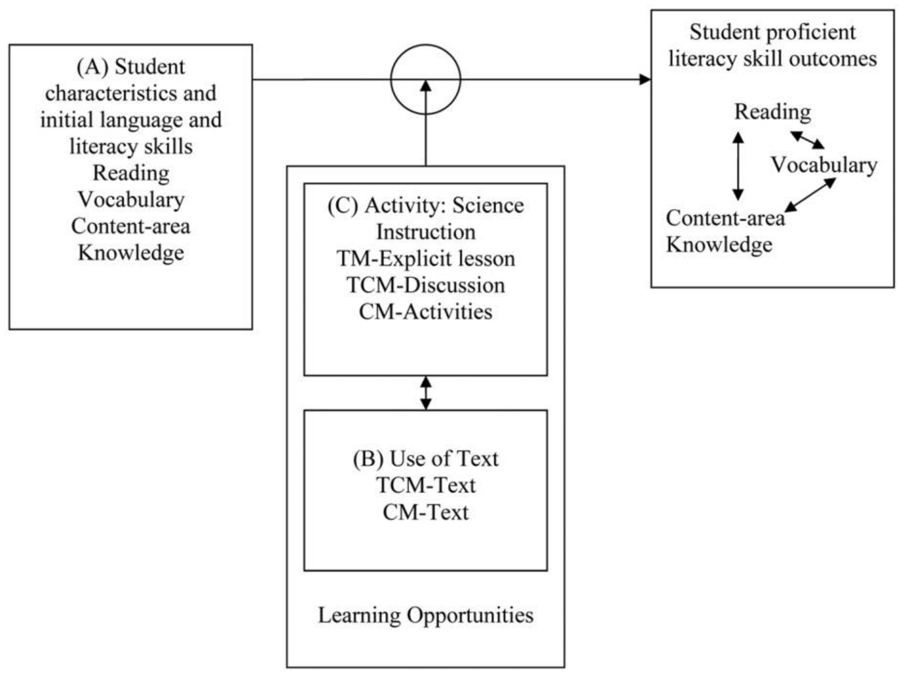


Figure 1. A model of science instruction and its effect on children's content-area literacy skill gains. Paths (arrows) show the hypothesized direction of effects. Note the circled interaction identifying the potential impact of child characteristic × instruction (CXI) interactions on students' literacy gains.

(Connor et al., 2011). Additionally, the effect of the learning opportunity on students' content-area literacy gains may depend on the characteristics and skills children bring to the classroom. This is designated in Figure 1 by the circled intersection between children's developing content-area literacy skills and the effect of the learning opportunity.

The National Academy of Science (2006) has called for vastly increased efforts in K–12 science education to prevent the United States from falling behind in the global economy. In line with these efforts, reform-based science instruction, which was observed in many of the participating classrooms, may also provide a particularly rich context for developing proficient content-area literacy. In reform-based instruction, students are taught to use their investigative skills to explore and actively seek answers to questions about specific science concepts (Duschl, Schweingruber, & Shouse, 2007; Von Secker, 2002). In this model of instruction, science classrooms provide rich contexts where students can exert their natural curiosity by investigating the world around them and engaging in language- and literacy-related activities to support these investigations (Duschl et al., 2007). When such investigative hands-on activities are integrated with concept-focused teaching, reading of science print materials, and journal writing, theoretically, both science understanding and reading achievement should be supported (Pearson, Moje, & Greenleaf, 2010; Romance & Vitale, 2001), but more research in this area is needed (National Academy of Science, 2006). In this study, we use a multidimensional approach to conceptualizing science instruction designed to capture key elements of the learning opportunities presented in Figure 1 (Connor, Morrison, et al., 2009). The approach is detailed in the Method section.

Why Second and Third Graders?

We selected second and third grade as the focus of inquiry for this study for a number of reasons. In second and third grade, children are still developing basic reading skills and the focus of instruction is transitioning from learning to read to reading to learn (Chall, 1996). We anticipate that early opportunities to gain content-area literacy skills might contribute to greater literacy proficiency overall. At the same time, we can examine whether science instruction might be systematically associated with students' gains in word-reading and reading-comprehension skills during this transition. As children get older, variability in word-reading skills, while still important, is less predictive of students' performance on, for example, state-mandated assessments than are vocabulary, comprehension, and content-area knowledge (Schatschneider et al., 2004).

Research Questions

The following research questions guided this exploratory investigation: (1) Does amount of time spent in specific science activities predict students' content-area knowledge, vocabulary, and reading gains? We hypothesize, based on the extant research, that specific science activities will predict students' literacy skill gains. (2) Is the association among specific types of science activities and students' outcome gains moderated by students' incoming reading, vocabulary, and content-area knowledge? We hypothesize that there will be child × science instruction interaction effects on literacy. For example, based on similar research for literacy (Connor, 2011), children with weaker content literacy skills in the fall may make greater gains when they engage in science instruction activities with the support of their teacher in contrast to when they are working independently or with peers. Following the same line of research, such activities might be associated with greater gains for students with stronger fall skills.

Method

Participants

The students and their second- and third-grade teachers in this study were part of a larger descriptive longitudinal study of students' literacy development from kindergarten through third grade at five schools (Connor, Morrison, & Petrella, 2004). Children proficient in English with no reported disabilities were eligible to participate in the study. Parents of students meeting these criteria were contacted about the voluntary participation of their children in the larger study. Consequently, for the original study, a total of 108 students were recruited in three waves at the beginning of their kindergarten or first-grade year and followed through second grade ($n = 88$) until they finished third grade ($n = 73$).

For this study, we focused only on second and third grade. Second- ($n = 33$) and third-grade ($n = 28$) teachers entered the study when one or more of the target

Table 1. Descriptive Statistics for Schools

School	% of Target Students Whose Parents Have High School Education or Less	Approximate Number of Classrooms per Grade	Number of Students School-wide	% of Students Qualifying for FRL (2001)	% of African American Students	% of Hispanic Students	% of White Students	Title I School?
School A	11	8	396	34	56	11	30	Yes
School B	5	7	371	28	39	5	47	No
School C	8	8	646	25	47	6	41	Yes
School D	0	4	409	25	35	5	52	No
School E	13	6	353	25	47	1	47	NA

Note.—Common Core of Data public school data 2000, 2001 (http://nces.ed.gov/ccd). Second-grade classrooms observed 1998 to 2001.

students were enrolled in their classrooms. Descriptive information is provided in Table 1 for the five schools and in Table 2 for the students who participated in the study. On average, children had cognitive scores falling within normal limits (IQ mean = 102, standard deviation 15) as measured by the Stanford-Binet Intelligence Scale-4 (Thorndike, Hagen, & Sattler, 1986). There were between one and six study children per classroom over the period of the study. With regard to attrition from second to third grade, data were missing for nine students: four students were missing classroom observation data and five students left the school district. Missing data analyses revealed no differences in family socioeconomic status (SES), IQ, gender, or second-grade fall scores. There were, however, significant differences in spring content knowledge (CK) and vocabulary (Voc) scores. Students missing in the third-grade sample showed significantly lower spring second-grade scores (CK raw score 51 vs. 40; Voc raw score 103 vs. 96). Teachers in all of the classrooms in this district met state certification requirements, including at least a bachelor's degree and a teaching certificate.

The participating school district was located in a large Midwestern city in which a highly selective university was located. Socioeconomically and ethnically diverse, the

Table 2. Descriptive Information about Student Characteristics and Outcomes (Raw Score)

	Mean	*SD*	Min	Max
Second grade:				
Parents' education (years)	16.22	3.20	3	23
Age in the fall of second grade (years; months)	7; 6	0; 6	7; 0	8; 0
Fall content-area knowledge	40.72	15.13	14.00	78.00
Fall word reading	49.50	15.43	23.00	86.00
Fall vocabulary	98.01	13.53	64.00	122.00
Spring content-area knowledge	48.75	17.60	11.00	86.00
Spring word reading	58.98	14.47	25.00	93.00
Spring vocabulary	101.66	14.49	67.00	140.0
Third grade:				
Fall content-area knowledge	55.49	15.96	27	86
Fall reading comprehension	61.49	12.63	26	89
Fall vocabulary	106.08	12.82	77	141
Spring content-area knowledge	63.30	15.31	28	84
Spring reading comprehension	66.23	12.95	34	94
Spring vocabulary	114.10	14.32	81	151
Spring word reading	71.11	14.36	27	96

district bordered a major metropolitan area and included neighborhoods of concentrated poverty. The five schools in which the study took place were fairly similar in the percentage of children who qualified for free or reduced-price lunch (FRL) but varied more in racial/ethnic composition. Two schools qualified for Title 1 funds. Based on parent educational levels as an indicator of SES, our target students were fairly representative of their school demographics, although we recruited fewer children from low-SES families in School D than might have been predicted based on school-wide SES levels. With regard to science curriculum, information reported by the research assistants who observed the classrooms indicated that no specific science curriculum was used. Teachers selected science activities and frequently relied on trade books for science text.

Measures

At the beginning of the study, students' parents completed background questionnaires, which were used to determine years of parent education (see Table 2). Children's academic/content-area knowledge, reading, and vocabulary were assessed in the fall and spring by trained graduate students. The target constructs were measured using selected subtests from the individually administered Peabody Individual Achievement Test—Revised (PIAT-R) (Markwardt, 1989). The PIAT-R consists of six subtests and has a reported median split-half reliability for the whole test of .98; test/retest reliabilities reported for the subtests are no lower than .90. Brief descriptions of each subtest used in this study are provided below. Raw scores, which were the number of items answered correctly, were used in the analyses.

Content-area knowledge. Content-area knowledge was assessed using the PIAT-R General Information subtest (Markwardt, 1989). For this task, the examiner read aloud increasingly difficult questions about general encyclopedic knowledge on a variety of topics from science, social studies, and literature. Forty percent of items 1 through 41, to which most of the participating children were exposed in the fall administration, incorporated questions about science topics. Fifty-eight percent of the items between 41 and 60, which represent additional items presented in the spring administration, were also science related.

Word reading. Word reading was assessed for second graders using the PIAT-R Reading Recognition subtest (Markwardt, 1989). In early parts of this task, children were asked to identify letters and letter sounds. For later items, students read increasingly unfamiliar words.

Reading comprehension. Reading comprehension was assessed for third graders using the PIAT-R Reading Comprehension subtest. In this task, children read a sentence or short passage and selected the most appropriate of four pictures. Eleven (31%) of the items within the range of students' responses (items 45–80) were expository in nature, with five of the 11 items specifically related to science concepts.

Vocabulary. Vocabulary was assessed using the Peabody Picture Vocabulary Test—Revised (PPVT-R) (Dunn & Dunn, 1981). Test/retest reliabilities range from the .70s to .90s; split-half reliabilities range between .60 and .80s. In this task, the examiner read increasingly unfamiliar words and the student had to select the most representative picture from a set of four.

Classroom Observations

Classrooms were observed for the entire school day, three times per year, one time each in the fall, winter, and spring. Second-grade classrooms ($n = 33$) were observed in 1997 to 2000 and third-grade classrooms ($n = 28$) were observed in 1998 to 2001. The coding system used in this study utilized a naturalistic approach (Connor et al., 2004). Rather than identifying the major types of instruction a priori, observations focused on recording the nature and content of the instructional activity in which most of the target children were engaged. Therefore, instead of trying to decide whether or not a particular activity was, for example, an experiment, the classroom observer described the activity including samples of teacher-child discourse using a written timed narrative transcript of the activities (transcripts available upon request from the first author). The goal was to have a narrative that was detailed enough to permit later coding but simple enough so that the observer could record activities as they occurred (the coding manual is available upon request from the first author), including both content-area instruction as well as noninstruction time (e.g., transitions, explaining procedures, and behavior management). In this way, the entire school day was recorded. Any activity that lasted at least 1 minute was described. Observers were rotated among classrooms to minimize the influence of observer expectations and, when possible, observations were scheduled on different days of the week. Interobserver agreement was obtained for all observers yearly by observing nonparticipating classrooms for a half day at the beginning of the school year. In every case, interobserver agreement was acceptable at levels of 95% or above for activity type and length in minutes.

The narratives were then coded for specific types of instructional activities. Activities were defined as instructional opportunities that had an identifiable instructional aim, for example, to teach children about sedimentary rocks. First, we coded for subject area—language arts, math, and science—and then for each type of activity occurring during the content-area instruction. Generic subactivities were identified, such as "discussion" and "teacher read aloud." These were then refined for the specific content area. Then, amount of time in each type of instructional activity was summed across the day for each teacher. Noninstruction time, including transitions, classroom organization, and disruptions, was coded separately and not included in the content-area amounts. Intercoder reliability among three coders was determined to be 92%.

To combine the coded subactivities, we identified two salient dimensions: (1) management and (2) delivery (see Table 3). The management dimension reflects the extent of interaction between the teacher and students, and helps to characterize

Table 3. Dimensions of Science Instruction

	Teacher-Managed (TM)	Teacher/Child-Managed (TCM)	Child/Peer-Managed (CPM)
Explicit lesson	Instruction		
Discussion/activities		Discussion Listening comprehension	Writing Science activity
Text reading		Teacher read aloud Student read aloud	Student sustained silent reading

the explicitness of the instruction, the scaffolding provided, or the encouraged independence of the students. Teacher/child-managed instruction, such as the teacher discussing with the class a science experiment just completed or leading children through a science experiment, is interactive and scaffolded. On the other hand, a teacher-managed classroom activity (e.g., the teacher explaining the water cycle) is less interactive. In both cases, however, the teacher is responsible for focusing the students' attention on the learning activity at hand. In child/peer-managed activities, the child is working independently or with peers without directly interacting with the teacher. For example, children might be working together in small groups to complete a science activity without the guidance of the teacher (Palincsar, Collins, Marano, & Magnusson, 2000). There is precedence for considering management as a dimension of science instruction. In a recently published meta-analysis on discovery learning, which included science among other content areas, the authors distinguished between unassisted discovery learning, where the students worked independently without scaffolding from the teacher, and enhanced discovery learning, where specific supports were provided by the teacher (Alfieri, Brooks, Aldrich, & Tenenbaum, 2010).

The second dimension, delivery, includes explicit lessons, discussion/activities, and text reading, and refers to how the content is delivered, the extent to which children are engaged in an activity during the instructional event, as well as the role of printed matter in the instructional activity (Yore, Bisanz, & Hand, 2003). During explicit lesson activities, science knowledge is clearly explained to children. For example, the teacher might be teaching a science concept, such as the changing lengths of daylight from summer to winter, using a chart. During discussion and activities, science knowledge is assumed to be more actively and jointly constructed between teachers and students and/or students with peers. The teacher and students discussing the types of habitats that earthworms require, and then the students and teacher together building a terrarium for earthworms, would be coded as discussion/activity. The principal difference between explicit lessons and discussion/activity is the explicit or interactive nature of the instruction, with explicit lessons more explicit and less interactive than discussion/activity events. While text might be a part of any activity, if the primary activity observed was text reading, it was defined as such. So, for example, if children were reading a book about rivers, this was coded text reading. However, if they were visiting a river, taking and comparing water samples, and writing a report of their results, this was coded discussion/activity.

These two dimensions operated simultaneously. Thus any observed activity was defined by both the management dimension and the delivery dimension (see Table 1). This yielded the five instructional variables used in this study: Teacher-Managed Explicit Lesson, Teacher/Child-Managed Discussion/Activities, Child/Peer-Managed Activities, Teacher/Child-Managed Text Reading, and Child/Peer-Managed Text Reading. The coded science activities were assigned to one of the five two-dimensional science instruction variables based on the judgment of the first two authors.

Second-grade instruction. Amounts of time spent on science instruction in second grade varied widely across classrooms and also across the year, ranging in the winter, for example, from 0 to 75 minutes. The mean daily amount of second-grade science instruction observed per teacher was 5 minutes per day in the fall, 15 minutes in the winter, and 13 minutes in the spring. When amounts of second-grade science instruction were summed across fall, winter, and spring observations for all teachers and divided by 3 to obtain the total mean number of minutes per day of science instruction, a mean amount of 11 minutes/day of science instruction across all three observations was obtained. However, this ranged from 0 to 38 minutes, so some teachers were providing fairly substantial amounts of science instruction during one school day. The means were relatively low because some teachers provided no science instruction during the observations.

On average, second-grade teachers spent 92 minutes per day in language arts instruction (range = 3 to 155 minutes/day), and this mean amount was relatively stable across the school year (fall = 95 min, winter = 98 min, spring = 83 min). There was a significant negative correlation between total amount of time in science and language arts instruction ($r = -.204$, $p < .05$). Thus teachers who spent more time teaching science tended to spend less time teaching language arts and vice versa. We found that, generally, the multidimensional variables captured the variability across classrooms in science instruction (see Table 4). As can be seen in Table 4, overall, second-grade teachers and students spent the greatest proportion of their time in science engaged in teacher/child-managed (TCM) discussion/activities. None of the multidimensional science instruction variables were significantly correlated with each other or to the total amount of language arts instruction provided.

Third grade. Slightly more time was observed in science instruction in third grade than in second grade, with, on average, 9 minutes observed in the fall, 16 minutes in the winter, and 16 minutes in the spring (see Table 4). The mean total amount of science instruction observed was 13 minutes ($SD = 11$), which ranged from 0 to 36 minutes; again, many teachers were observed to provide substantial amounts of science instruction, whereas others provided none. Of note, all but 17 third graders in the sample were observed to receive science instruction sometime throughout the school year. Similar to second grade, generally, more time was spent in TCM discussion/activities than in the other activities (see Table 4). Amounts of science and language arts instruction were not significantly correlated; however, third-grade teachers who provided more time in child/peer-managed (CPM) discussion/activi-

Table 4. Science Instruction Descriptives (Minutes per Day), Second Grade and Third Grade

	Mean	*SD*	Min	Max	% of Time
Second grade:					
TM explicit lesson	.93	2.97	.00	16.33	8
TCM discussion/activities	5.82	5.43	.00	19.67	53
CPM discussion/activities	2.02	5.03	.00	25.00	19
TCM text reading	1.62	3.31	.00	17.00	15
CPM text reading	.56	1.94	.00	10.33	5
Total science	10.95	9.75	.00	38.33	
Total language arts	92.39	25.94	52.67	154.67	
Third grade:					
TM explicit lesson	1.39	2.49	.00	10.00	11
TCM discussion/activities	4.08	4.23	.00	14.00	32
CPM discussion/activities	2.74	3.95	.00	16.33	22
TCM text reading	3.35	4.93	.00	17.33	26
CPM text reading	1.17	2.30	.00	10.33	9
Total science	12.73	11.09	.00	35.67	
Total language arts	137.54	44.65	57.67	250.67	

Note.—TM = teacher-managed, TCM = teacher/child-managed, CPM = child/peer-managed.

ties were also much more likely to provide more time in CPM text-reading activities ($r = .81$, $p < .001$). Those teachers who provided more time in teacher-managed (TM) explicit lesson activities were more likely to provide more time in TCM discussion/activities ($r = .69$, $p < .001$).

Analytic Strategies

We designed our models to test the theoretical framework displayed in Figure 1. To do this, we modeled the outcome constructs—reading, vocabulary, and content-area knowledge—as a function of (*a*) children's skills in the fall and (*b*) the instruction they received, and then (*c*) tested whether children's skills in the fall moderated the effect of the instruction they received by testing fall skill × instruction type interactions. Because the study was exploratory in nature and because we hypothesized that there would be CXI interactions, we tested all of the possible interactions and then trimmed the interactions that did not significantly contribute to our outcomes.[1]

Data had a nested structure, with children nested in classrooms. Thus hierarchical linear models (HLM; Raudenbush & Bryk, 2002) were used, which can accommodate this data structure and allow us to test cross-level child × classroom interactions. Additionally, our science instruction variables were skewed because in some classrooms no science instruction was observed and HLM is robust in this regard. Models were built systematically. We began with an unconditional model for each outcome. Child variables were then added to the model. Variables for which random effects were not significantly greater than zero were fixed. Those variables that did not significantly predict child outcomes were trimmed from the model to preserve parsimony. Classroom instruction variables were added at level 2 and then CXI interactions were tested. Highly correlated variables (see Table 5) were not included in the same model, and decisions to include variables were made theoretically.[2]

Table 5. Correlations among Child Variables for Second and Third Grade

	1	2	3	4	5	6
Second grade:						
1. Home literacy environment score	–					
2. Fall content-area knowledge raw score	.515***	–				
3. Fall word-reading raw score	.212	.488***	–			
4. Fall vocabulary raw score	.679***	.738***	.415***	–		
5. Spring content-area knowledge raw score	.489***	.820***	.494***	.747***	–	
6. Spring word-reading raw score	.269*	.501***	.854***	.405***	.518***	–
7. Spring vocabulary raw score	.576***	.747***	.554***	.816***	.744***	.549***
Third grade:						
1. Home literacy environment score	–					
2. Fall content-area knowledge raw score	.500***	–				
3. Fall reading comprehension raw score	.321**	.599***	–			
4. Fall vocabulary raw score	.573***	.724***	.682***	–		
5. Spring content-area knowledge raw score	.470***	.828***	.662***	.760***	–	
6. Spring reading comprehension raw score	.372**	.625***	.811***	.673***	.662***	–
7. Spring vocabulary raw score	.529***	.757***	.670***	.876***	.739***	.684***

$^{*}p < .05$.
$^{**}p < .01$.
$^{***}p < .001$.

Results

Overall, students' content-area literacy skills, specifically content knowledge, word reading (grade 2), reading comprehension (grade 3), and vocabulary, improved from fall to spring in each grade and, for content knowledge and vocabulary, improved from second to third grade (see Table 2). Individual models were run for each grade. Hence we first provide the second-grade results and then the third-grade results for each research question separately. The proportion of variance explained between classrooms, or the intraclass correlations (ICCs), increased for spring scores when fall score was included in the model, which suggests that although the classroom environment accounted for very little variance in students' spring scores, classrooms accounted for more variance with regard to students' residualized gains. At the same time, in general, the relatively small ICCs, with the exception of residualized gains for vocabulary in third grade, indicated that there were greater differences in achievement among children who shared classrooms than between classrooms. ICCs in second grade (residualized gain) were .111, .068, and .058 for content knowledge, reading, and vocabulary, respectively; in third grade, ICCs were .005, .001, and .497, respectively.

Does Time Spent in Science Instruction Predict Content-Area Knowledge, Vocabulary, and Reading Gains? Are There CXI Interactions?

Second grade.

Content knowledge. The final model explained 67% of the variance in children's spring content-area knowledge scores and explained all of the classroom-level variance (see Table 6). In general, our hypotheses were supported, with some notable exceptions. Overall, observed science instruction affected students' content knowledge gains (residualized change). Specifically, TCM text reading had a generally negative effect on students' gains, and CPM text-reading activities had a generally positive effect on students' gains. As hypothesized, there were CXI interactions for two types of instruction: CPM activities and CPM text reading. Whereas students generally showed greater gains in content-area knowledge when they were in classrooms that spent more time in CPM activities, this association held only for students who began the school year with average to stronger fall content knowledge scores, modeled at the 75th and 50th percentiles, respectively (see Fig. 2, top). In contrast, for students who started second grade with weaker content knowledge scores, modeled at or below the 25th percentile, in general, the more time spent in CPM activities, the weaker were their gains in content knowledge. Keep in mind that fall content-area knowledge and the other outcome variables are continuous variables. Therefore, for the figures, we conservatively modeled plausible scores at the 25th, 50th, and 75th percentiles, which fell well within sample ranges. These translate to fall scores of 29, 41, and 51, respectively, for fall content knowledge scores. We could have chosen more extreme high scores (e.g., raw score = 75) and shown CPM activities to have an even stronger positive association with content knowledge gains for children whose fall content-area knowledge scores fell at the 90th percentile. In the models, all variables not represented in the figures were held constant at their respective sample grand mean.

Students in classrooms where more time was spent in CPM text-reading activities demonstrated greater content-area knowledge gains on average than did children in

Table 6. Results Examining the Impact of Child and Instruction Variables on Second-Grade Students' Spring Content Knowledge (CK), Word Reading (WR), and Vocabulary (Voc)

Fixed Effect	CK	WR	Voc
Intercept (mean), γ_{00}	49.36***	57.44***	101.96***
Child-level variables:			
Fall CK, γ_{10}	.98***	.07	.22**
Fall WR, γ_{20}		.70***	.18***
Fall Voc, γ_{30}			.61***
Classroom-level variables:			
TCM text reading effect, γ_{01}	−.38*	.01	−.19+
CPM text reading effect, γ_{02}	1.65***	.77**	−.72**
CPM activities effect, γ_{03}	.002	.09	−.15+
Total language arts, γ_{04}	−.007	.0003	−.06*
TCM discussion effect, γ_{05}	−.015	.05	.35**
TM explicit lesson effect, γ_{06}	.189	−1.50***	.26*
Child × classroom-level interactions:			
Fall CK interaction × CPM activities interaction, γ_{11}	.01***	−.02***	
Fall CK × CPM text interaction, γ_{12}	.05*		
Fall CK × TM explicit lesson interaction, γ_{13}		−.16***	
Fall CK × TCM text reading interaction, γ_{14}			.02***
Final estimation of variance components:			
Random effect variance:			
Intercept 1, U_0	11.78	.028	.14
Level 1, R	89.17	43.40	48.28
Unconditional ICC	.0007	.001	.010

Note.—All continuous variables are grand mean centered. TM = teacher-managed, TCM = teacher/child-managed, CPM = child/peer-managed.

$^{+}p < .10$.

$^{*}p < .05$.

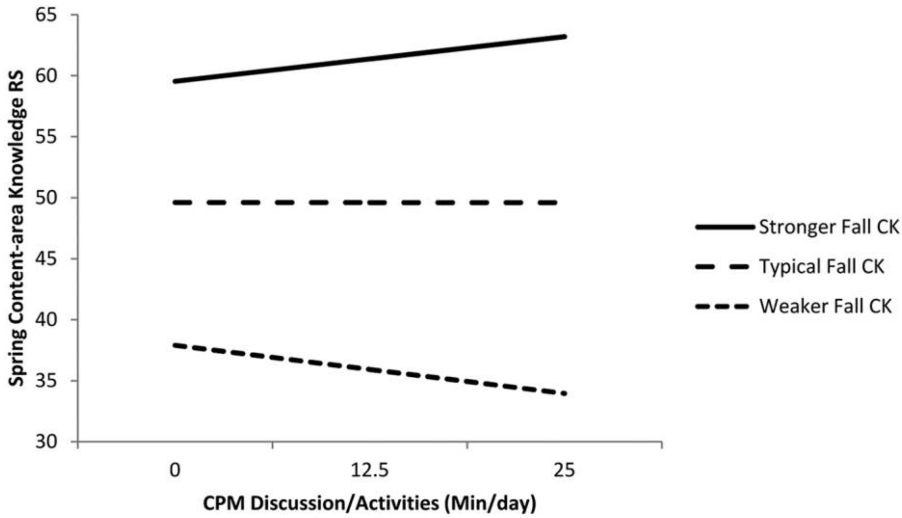
$^{**}p < .01$.

$^{***}p < .001$.

classrooms where less time was spent in CPM text-reading activities. This effect was somewhat greater for students who began second grade with stronger content knowledge scores.

Word reading. Results for the final model for second-grade spring word reading, which explained 79% of the variance in children's spring word-reading scores, are provided in Table 6. Children's fall word skills positively predicted spring word-reading skills. Overall, the fall vocabulary score variable did not significantly predict word skills and so was trimmed from the model. Again, our hypotheses were largely supported. Students in classrooms that spent more time in CPM text-science reading activities generally demonstrated greater gains in word-reading skills than did students in classrooms that spent less time in these activities. For every additional minute children spent in CPM text activities, their word-reading scores were predicted to increase 0.77 points. Total amount of language arts instruction was not associated students' word-reading skills gains.

There were also CXI interactions (see bottom of Fig. 2). An unanticipated finding was that children with lower fall content-area knowledge scores (below about the 50th percentile of the sample) demonstrated greater word-reading gains when in classrooms that offered greater amounts of time in CPM activities than did children with similar content-area knowledge scores in classrooms with lesser amounts of time spent in CPM activities. At the same time, children with stronger fall content-



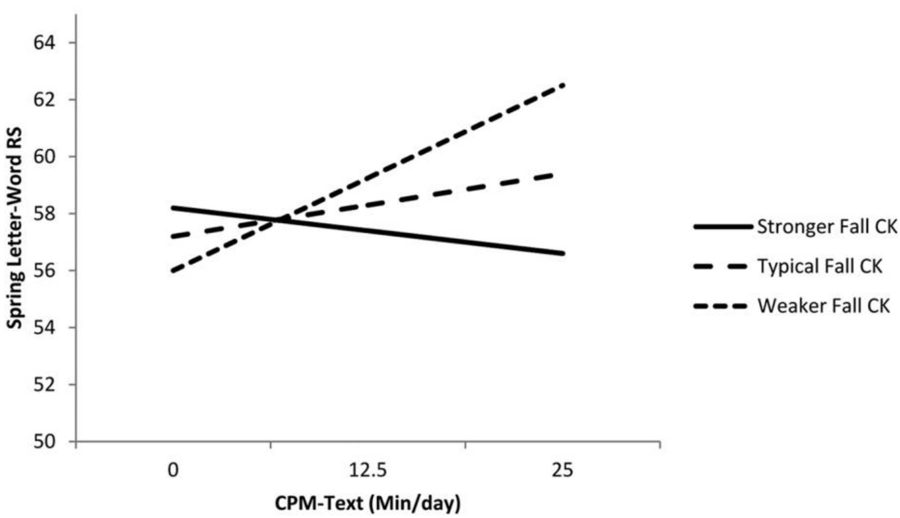


Figure 2. Second-grade spring content-area knowledge (CK) raw score predicted by child/peer-managed (CPM) discussion/activities (top) and word reading predicted by CPM text reading (bottom). Higher fall CK is fitted at the 75th percentile of the sample, average CK is fitted at the 50th percentile, and lower CK is fitted at the 25th percentile of the sample.

area knowledge scores (falling above the 75th) tended to demonstrate weaker word-reading score gains when they were in classrooms that spent more time in CPM activities compared to children with similar scores in classrooms with less time spent in such activities. This can be seen in Figure 2 (bottom), which shows spring word-reading score varying as a function of changes in the amount of CPM activities for children with fall content-area knowledge scores modeled at the 25th, 50th, and 75th percentiles of the sample.

There was also a fall content-area knowledge × TM explicit lesson interaction. This is interpreted in the same way as for the CPM activities interaction with fall content-area knowledge. Essentially, students in classrooms with more rather than less time spent in TM explicit lesson activities generally demonstrated weaker word-reading gains by spring. If children started second grade with weaker content-area knowledge, the amount of classroom TM explicit lesson activities had a positive effect on their spring word-reading gains.

Vocabulary. Second graders' vocabulary, word-reading, and content-area knowledge scores in the fall positively predicted their vocabulary score in the spring (see Table 6). On average, students in classrooms that spent more time in language arts instruction demonstrated weaker gains in vocabulary compared to students in classrooms that spent less time in language arts. With regard to science instruction, students who attended classrooms with greater amounts of time observed in TCM discussion/activities and TM explicit lesson activities demonstrated greater vocabulary gains than did students who attended classrooms with less time in these activities. In contrast, children in classrooms where more time was spent in CPM text reading activities tended to demonstrate weaker vocabulary gains than did students in classrooms where less time was spent in CPM text reading activities.

Supporting our hypothesis, there was also a significant CXI interaction between students' fall content-area knowledge scores and time spent in TCM text reading. Students in classrooms where more time was spent in TCM text reading demonstrated weaker vocabulary gains compared to students in classrooms with less time spent in TCM text reading. This negative effect was greater for students with stronger fall content-area knowledge scores. For students with lower fall content-area knowledge scores, TCM text reading had no systematic association with students' gains. All results controlled for students' fall vocabulary, general knowledge, reading recognition scores, and total amount of science and language arts instruction. The final model explained 79% of the variance in children's spring vocabulary scores.

Third grade.

Content knowledge. Fall content knowledge and reading comprehension predicted third graders' spring content knowledge scores (see Table 7; variance explained = 75%), and, supporting our hypotheses, there were CXI interactions for CPM activities, TCM discussion/activities, and TM explicit lessons. Figure 3 (top) reveals patterns of results that were highly similar to those found in second grade for CPM activities. In general, students in classrooms with more time spent in CPM activities demonstrated greater gains in content knowledge as long as they began third grade with typical or above average content knowledge fall scores. CPM activities had a negative effect on content knowledge gains for students who began the year with weaker scores. In contrast, TCM discussion/activities had a generally positive impact on students' gains for students with weaker and more typical fall content knowledge. Moreover, the effect was greater for students with weaker fall content knowledge (see Fig. 3, middle). Overall, children in classrooms with greater amounts of TM explicit lesson activities showed weaker gains in content knowledge, except for students with stronger fall skills. For them, there was a positive effect on content knowledge gains.

Reading comprehension. The final model for reading comprehension revealed that, in general, the more time third graders were observed reading science text (CPM text reading), the greater were their reading comprehension gains, except for students with weaker fall reading comprehension scores (see Table 7 and Fig. 3, bottom). For them there was a small but negative association. It should be noted that the amount of time in language arts instruction had no significant association with reading comprehension gains.

Table 7. Results Examining the Impact of Child and Instruction Variables on Third-Grade Students' Spring Content Knowledge (CK), Reading Comprehension (RC), and Vocabulary (Voc)

Fixed Effect	CK	RC	Voc
Intercept (mean), γ_{00}	63.79***	66.52***	115.42***
Child-level variables:			
Fall CK, γ_{10}	.59***	.18***	
Fall RC, γ_{20}	.34**	.69***	.14*
Fall Voc, γ_{30}			.90***
Classroom-level variables:			
TCM text reading effect, γ_{01}	.02	−.03	.11
CPM text reading effect, γ_{02}		.26	−.14
CPM activities effect, γ_{03}	−.15		
Total language arts, γ_{04}	−.009	−.01	−.04*
TCM discussion effect, γ_{05}	.55+	−.28	.23
TM explicit lesson effect, γ_{06}	−.87*	.18	−.33
Child × classroom-level interactions:			
Fall CK interaction × CPM activities, γ_{11}	.06*		
Fall CK × CPM text, γ_{12}		−.03+	
Fall CK × TM explicit lesson, γ_{13}	.62***		
Fall RC × CPM text, γ_{14}		.02**	
Fall CK × TCM discussion, γ_{15}	−.04**		
Fall Voc × total language arts, γ_{16}			−.002+
Final estimation of variance components:			
Random effect variance:			
Intercept 1, U_0	.05	1.03	26.27***
Level 1, R	58.56	56.81	23.81
Unconditional ICC	.009	.0004	.058

Note.—All continuous variables are grand mean centered. TM = teacher-managed, TCM = teacher/child-managed, CPM = child/peer-managed.

$^{+}p < .10$.

$^{*}p < .05$.

$^{**}p < .01$.

$^{***}p < .001$.

Vocabulary. The final model for third-grade vocabulary revealed that science instruction had no significant effect on third graders' vocabulary gains (see Table 7). Total time in language arts was negatively associated with vocabulary gains.

Discussion

Findings reveal that learning opportunities during science instruction are generally associated with second- and third-grade students' gains in components of content-area literacy: content knowledge, reading, and vocabulary. Overall, our hypotheses informed by the model in Figure 1 were largely supported, albeit with greater complexity than anticipated. Aspects of science instruction, which in the model is the learning opportunity, were positively and negatively associated with significant gains in all three of the outcomes in both second and third grade. Moreover, as hypothesized, the association of certain types of science instruction with students' literacy gains generally depended on students' initial reading, vocabulary, and content-area knowledge. The patterns of associations and CXI interactions differed somewhat for second and third graders, but there were similarities as well, particularly with regard to content knowledge outcomes. Thus, science instruction may be more effective

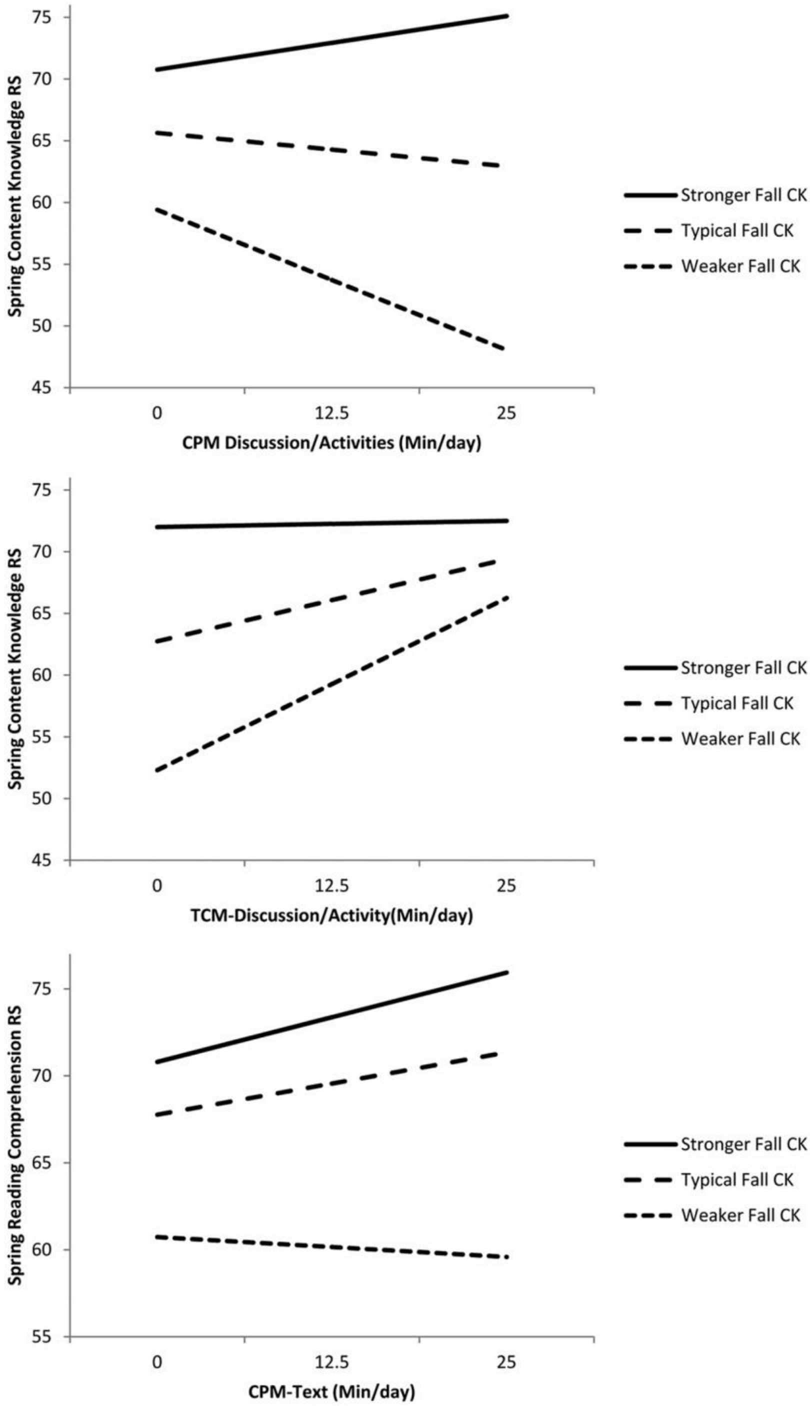


Figure 3. Third-grade content-area knowledge gains as a function of amount of CPM discussion/ activities (top) and TCM discussion/activities (middle) by fall content-area knowledge (CK) score interaction. Lower CK is fitted at the 25th percentile for the sample in the fall, average CK is fitted at the 50th percentile, and higher CK is fitted at the 75th percentile of the sample. Third-grade reading comprehension (RC) gains as a function of CPM text and fall RC (bottom), where lower RC is modeled at the 25th percentile of the sample, typical at the 50th, and higher fall RC at the 75th percentile.

when it takes individual child differences into consideration. At the same time, the results illustrate the potential promise of content-area instruction as a strategy to promote children's development of proficient literacy skills and content knowledge more generally.

We offer a few words of caution before we discuss our results further. First, we relied on the naturally occurring variability in teachers' practices and students' skills to identify associations among instruction amounts and types observed in classrooms and students' outcome gains. For this reason, the results of this study should be interpreted conservatively. There are other plausible reasons for the associations found in the models. For example, teachers' warmth and responsiveness are consistently associated with stronger student outcomes (NICHD, Early Child Care Research Network, 2004). It may be that teachers who are generally more responsive to their students are more likely to provide opportunities for students to participate in TCM discussion/activities or may be more effective in general throughout the school day. In addition, we did not make any judgments regarding the quality of the science instruction provided, although it is clear from the extant research that the quality of science education is important to consider (Duschl et al., 2007). Disambiguating these potential sources of influence on students' achievement was beyond the scope of this study but is clearly worthy of additional study.

This was a secondary analysis of an existing data set and, as is frequently the case when a study is not designed for the specific research questions, information that might be desired was not available. The larger study was intended to examine the effect of instruction on students' literacy skill gains and was not designed with specific questions about science outcomes in mind. As a result, no specific science knowledge assessments were conducted, which would have strengthened the current investigation. Our measure of content-area knowledge did rely heavily on science content. Indeed, about half of the questions administered to the children in the sample, based on raw scores obtained, were about science content. Thus, in order for students to do well on our content-area knowledge assessment, a good grasp of science content was likely required. At the same time, for science instruction to be associated with the more general constructs of knowledge tested on the PIAT General Knowledge Test indicates that science instruction may have a powerful influence on students' learning beyond just science knowledge. Additionally, there was no assessment of student comprehension in second grade. Although we might infer that students with strong content knowledge, vocabulary, and word-reading skills would also tend to have strong comprehension skills (Snow, 2001), and it was clear from the third-grade correlations that this was the case, a measure of comprehension in second grade would have strengthened the current investigation.

Associations among Types of Science Instruction and Students' Content Literacy Outcomes

There is a wealth of evidence that background knowledge and vocabulary are highly predictive of reading proficiency (NICHD, 2000), and our results provide promise that science instruction may support the development of these key components. However, the patterns of associations were complex, and below we highlight findings that may be particularly useful for informing practice and future research. An overview of the complete results by type of instruction and outcome for second and third grade is provided in Table 8. By general standards, effect sizes were small for second and moderate for third grade (Hill, Bloome, Black, & Lipsey, 2008).

In general, reading science text, either by the teacher or by students, was associated with gains in content-area literacy although there were child × instruction interactions (see Table 8). Especially for students with weaker initial skills, reading and discussing expository text may support key components of literacy, including vocabulary and content-area knowledge, which would tend to support reading skills in general (Guthrie et al., 2004; Williams et al., 2005).

We found that second graders in classrooms with more opportunities to participate in science lessons that were explicitly taught by their teacher along with opportunities to discuss and work with their teachers in the context of science activities demonstrated stronger vocabulary gains than did second graders in classrooms with fewer such opportunities. It may be that during these kinds of lessons, science content and science vocabulary are introduced and explained. This is likely to build vocabulary as well as content-area knowledge generally (Biemiller & Boote, 2006). This finding is particularly encouraging because there are few identified instructional strategies that change the rate of children's vocabulary growth (Hart & Risley, 1995), particularly on standardized measures. Further, there is experimental evidence that, generally, schooling has little effect on students' vocabulary growth (Morrison, Smith, & Dow-Ehrensberger, 1995). Some specific targeted strategies appear to be related to learning specific vocabulary (Beck, McKeown, & Kucan, 2002; NICHD, 2000), but it is not clear that the rate of general vocabulary growth overall is improved. In contrast, our results suggest that exposure to more explicit, teacher-provided science instruction and interactive discussions among teachers and students in the context of science may be contributing to students' vocabulary growth.

These two types of science instruction were also associated with greater gains in content knowledge for third graders, but there were CXI interactions. The amount of time third graders participated in discussions and science activities with their teacher was associated with stronger outcomes only for third graders with weaker fall content knowledge. For more explicit science instruction, the association was reversed. More time in teacher-managed explicit science lessons had a greater positive effect for third graders with stronger fall content knowledge. It may be that the more interactive nature of the teacher-scaffolded discussions and experiments, as compared to the more didactic teacher-managed lessons, better supported learning for students with weaker content knowledge. These results also support findings from a recent meta-analysis that, in general, discovery learning opportunities that are scaffolded (i.e., teacher/child managed) generally have a greater positive impact on students' learning than more direct (i.e., teacher managed) instruction, which in turn have a greater positive impact than unassisted discovery learning (i.e., child and peer managed) (Alfieri et al., 2010).

Hands-on and other science activities in which students are involved in sense making about a phenomenon, including writing activities, are considered to be an important part of high-quality science education (Duschl et al., 2007). Much of science education research starts with the premise that hands-on and exploration activities should be more strongly associated with greater science knowledge gains than the other types of science instruction. However, these studies did not consider individual child differences. The results of this study revealed that the effect of such instruction depended on students' fall content-area knowledge in both second and

Table 8. Overview of the Association between Types of Second- and Third-Grade Instruction Provided and Students' Residualized Change (Gains) in Scores

Type of Instruction and Grade	TM Explicit Lesson		TCM Discussion/ Activities		CPM Discussion/ Activities		TCM Text Reading		CPM Text Reading		Language Arts	
	Higher CK	Lower CK	Higher CK	Lower CK	Higher CK	Lower CK	Higher CK	Lower CK	Higher CK	Lower CK	Higher Voc	Lower Voc
Content-area knowledge (CK):												
Second					+	− −	−.30	−.30	++	+		
Third	+	− −	0	++	++	−	+	0	na	na		
Vocabulary (Voc):												
Second	.12		.29				− −	+			−.24	−.24
Third									na		−.36	−.36
Word reading (WR) in second grade and reading comprehension (RC) in third grade:												
									Higher R	Lower R		
Second WR	− −	+			−	++			.23	.23		
Third RC									++	−		

Note.—Reported effect sizes (*d*) are for a 1 *SD* change, which is computed by multiplying the coefficient in the model by 1 *SD* (see Table 3) and dividing by the student level *SD* (the square root of the variance, *R*) in the model. Only statistically significant effects are provided. For child × instruction interactions, effect sizes can be larger or smaller depending on the modeled values. For example, in Figures 2 and 3, fall scores are modeled at the 25th, 50th, and 75th percentiles of the sample for the entire range of time spent in the instructional activity. Plus (+) indicates a generally positive effect size with more +'s indicating a larger effect. Minus (−) indicates a negative effect size, and 0 indicates no or a negligible effect size.

third grade. Students in classrooms where they were observed to be conducting their own experiments and working with peers generally showed greater gains in content knowledge than did students in classrooms where less time in such activities was observed. However, this was the case only for students with stronger fall content knowledge scores. Students with weaker fall skills showed weaker gains (see Fig. 2, top, and Fig. 3, top). Hence, the meta-analysis by Alfieri and colleagues (2010) may have underestimated the impact of unassisted discovery learning for students with stronger science knowledge and overestimated the impact for students with weaker knowledge.

The observed child × instruction interaction for child/peer managed activities has implications for students who attend schools in higher poverty neighborhoods. In the study reported here, children with weaker fall vocabulary and content-area knowledge scores were much more likely to have parents with less education and to have lower home literacy scores (see Table 5). This association has been well documented (Hart & Risley, 1995). Thus, providing students who have weaker vocabulary and content knowledge more time in student- and peer-managed science activities than in teacher-scaffolded science activities might possibly limit their learning gains. This would then potentially contribute to well-documented achievement gaps (Jencks & Phillips, 1998). There is corroborating evidence in another study comparing traditional and reform-based curricula that a reform-based curriculum may be less effective for children attending higher poverty schools if such instruction is not well scaffolded by the teacher at the outset (Lee & Luykx, 2006). In another study, students attending higher poverty schools generally made greater gains in inquiry lessons than in traditional science curricula, although they did not achieve the levels of students in more affluent schools (Blanchard et al., 2010). Notably, Blanchard, Southerland, and colleagues involved students in well-scaffolded guided inquiry, which is similar to the teacher/child-managed activities described in this study. Children with learning disabilities are also likely to have weaker vocabulary skills, and there is also some evidence that they may be less well served by guided inquiry when heterogeneous groupings are used (Palincsar et al., 2000). Our findings support the growing body of literature that suggests that implementation of science education reform should be mindful of the knowledge and skill sets students bring with them into the classroom (Lee & Luykx, 2006).

Taken together, these results suggest that a one-size-fits-all approach to science instruction might not yield stronger outcomes for some children. Given the historical centrality of activity-based instructional approaches to science education, these results indicate that experimental research investigating the influence of these kinds of activities on students' science and literacy learning, and potential child × instruction interactions, would be informative (Marx & Harris, 2006).

Implications

In the current policy climate, and as more states introduce state-mandated testing of science achievement and the Common Core Standards, these findings have potential implications for research and practice. First, schools are going to be held increasingly accountable for students' outcomes on science achievement tests (Marx & Harris, 2006). Thus, finding ways to improve students' science literacy will become progressively more important. These results suggest that as early as second grade, a focus on content-area literacy is appropriate. The activities that comprise science instruction were systematically associated with critical components of proficient literacy, depending on students' initial skills.

Second, we observed that teachers provided significantly more instructional time in language arts than they did in science, with greater differences between the two than has been reported in other studies (e.g., Fulp, 2002). The results of this study suggest that time spent in science does not detract from supporting students' literacy skill gains. Instead, science may offer the opportunity to read expository text, learn new concepts, increase vocabulary, practice reading comprehension strategies, and build content knowledge (Krajcik & Sutherland, 2010; Pearson et al., 2010; Van den Broek, 2010).

Returning to Figure 1, in our model we predicted that children's entering content knowledge, vocabulary, and reading skills would moderate the association between the instruction they received and their achievement gains, which we call child $\times$ instruction (CXI) interactions. The observed CXI interactions indicate that what is effective science instruction for one student may be ineffective at different grades or for another student in the same grade but with a different profile of content-area knowledge and literacy skills. Thus, science instruction that is integrated with literacy instruction, informed by valid and reliable formative assessment of students' skills, and individualized or differentiated in line with individual child differences may be more effective in supporting students' content-area knowledge and literacy learning than science instruction that does not take into account individual child differences. Recent attempts to differentiate science instruction and integrate this instruction with explicit comprehension instruction appear to be promising (Cervetti, Damico, & Pearson, 2006; Connor et al., 2010; Pearson et al., 2010) but are not conclusive. Moreover, without reliable and valid assessment of key skills, it is difficult to differentiate science instruction, and such assessments are not always readily available to educators. Nevertheless, educators may be able to design more effective science instruction by intentionally incorporating the wealth of accumulating evidence on effective literacy instruction and what is known about individual student differences, together with evidence on effective science instruction. Combining this with the careful assessment of students' content-area knowledge, language, and literacy skills, we may be able to design and implement more optimal science instruction that will build content-area knowledge and proficient reading skills for each student in the classroom.

Notes

The authors wish to thank Dr. Ana Thompson for her work on an early version of this manuscript; Dr. Laura Lang, Dr. Alysia Roehrig, Dr. Jeanine Turner, Dr. Sharilyn Steadman, and Dr. Stephanie Al Otaiba for comments on a version of this manuscript; and Jessica Toste, Melissa Luck, Novell Tani, and Meghann Montgomery for feedback on the revisions. We are deeply indebted to the school district administration, teachers, parents, and students who participated in the study and who made this research possible. We also thank the Loyola, the UM Pathways to Literacy, and the FSU/FCRR/FCR-STEM Individualizing Student Instruction Project Research Teams. Funding was provided by the U.S. Department of Education, Institute for Education Sciences, Cognition and Student Learning (R305H04013); the National Institute of Child Health and Human Development (R01 HD48539 and R01 HD27176); and the Florida Department of Education through a grant awarded to the Florida Center for Research in Science, Technology, Engineering and Mathematics

at Florida State University (FCR-STEM). The opinions expressed are those of the authors and do not represent views of the funding agencies.

1. Level 1 model: Y_{ij} (predicted spring score) $= \beta_{0j} + \beta_{1j}$ *(fall score$_{ij}$) $+ \beta_{2j}$ *(fall covariate score$_{ij}$) $+ r_{ij}$. Level 2 model: $\beta_{0j} = \gamma_{00} + \gamma_{01\ldots n\text{-}1}$*(classroom instruction variables [e.g., TM explicit lesson, CM activities, total language arts]$_j$) $+ u_{0j}$, $\beta_{1j} = \gamma_{10} + \gamma_{11}$*(type instruction$_j$), and $\beta_{2j} = \gamma_{20} + \gamma_{21}$ *(type instruction$_j$). In this model, Y_{ij} is the predicted spring score (e.g., content-area knowledge) for child i in classroom j. All continuous variables are centered at their sample grand mean. The intercept coefficient, γ_{00}, represents the fitted mean spring score for the sample, holding all other variables constant at their sample means. In other words, the intercept represents the predicted or fitted score for a child who begins the school year with average fall scores and who attends a classroom where the average amount of each instruction type is provided. $\gamma_{01\ \ldots\ n\text{-}1}$ represent the coefficients for each of the classroom instruction types, and γ_{11} and γ_{21} represent the child score × instruction type interactions. The proportion of variance explained by the final model was computed by subtracting the total variance in the final model from the total variance in the unconditional model and dividing by the total variance in the unconditional model. The intraclass correlation (ICC), which represents the proportion of the variance in the outcome falling between classrooms, is computed using the unconditional model results by dividing the level 2 between-classroom variance (u_0 or τ) by the sum of the level 2 between-classroom variance and the level 1 within-classroom variance (r or σ^2). Model building was systematic.

2. For example, in second and third grade, fall content knowledge and vocabulary were not included in the same model and the autoregressor was used. In third grade, amounts of CPM text reading and CPM discussion/activity were highly correlated ($r = .813$, $p < .001$). Therefore, time spent in CPM discussion/activity instruction was included when the outcome of interest was content knowledge and vocabulary, whereas CPM text reading was included when the outcome was reading comprehension.

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2010). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*. Advance online publication, doi: 10.1037/a0021017

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction; solving problems in the teaching of literacy*. 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.

Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability? Investigating the relative effectiveness of guided inquiry and traditional, didactic laboratory instruction. *Science Education*, **94**(4), 577–616.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (6th ed., Vol. **1**, pp. 793–828). Hoboken, NJ: Wiley.

Cervetti, G., Damico, J., & Pearson, P. D. (2006). Multiple literacy, new literacies, and teacher education. *Theory into Practice*, **45**(4), 378–386.

Chall, J. S. (1996). *Stages of reading development* (2nd ed.). Orlando, FL: Harcourt Brace.

Connor, C. M. (2011). Child by instruction interactions: Language and literacy connections. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook on early literacy* (3rd ed., pp. 256–275). New York: Guilford.

Connor, C. M., Kaya, S., Luck, M., Toste, J., Canto, A., Rice, D. C., . . . Underwood, P. (2010). Content-area literacy: Individualizing student instruction in second grade science. *Reading Teacher*, **63**(6), 474–485.

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

Connor, C. M., Morrison, F. J., Fishman, B. J., Schatschneider, C., & Underwood, P. (2007). The early years: 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 by instruction interactions. *Journal of Educational Psychology*, **96**(4), 682–698.

Connor, C. M., Morrison, F. J., Schatschneider, C., Toste, J., Lundblom, E. G., Crowe, E., & Fishman, B. (2011). Effective classroom instruction: Implications of child characteristic by instruction interactions on first graders' word reading achievement. *Journal for Research on Educational Effectiveness*, **4**(3), 173–207.

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., . . . Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of child by instruction interactions on first graders' literacy development. *Child Development*, **80**(1), 77–100.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: American Guidance Service.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *National Research Council, taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.

Fulp, S. L. (2002). *2000 National Survey of Science and Mathematics Education: Status of elementary school science teaching*. Chapel Hill, NC: Horizon Research.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., . . . Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, **94**(3), 403–423.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes.

Hill, C., Bloome, H., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, **2**(3), 172–177.

Jencks, C., & Phillips, M. (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution

Juel, C., & Minden-Cupp, C. (1998). *Learning to read words: Linguistic units and strategies*. Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.

Krajcik, J. S., & Sutherland, L. M. (2010). Supporting students in developing literacy in science. *Science*, **328**, 456–459.

Lee, O., & Luykx, A. (2006). *Science education and student diversity*. New York: Cambridge University Press.

Markwardt, F. C. (1989). *Peabody Individual Achievement Test—revised*. Circle Pines, MN: American Guidance Service.

Marx, R. W., & Harris, C. J. (2006). No Child Left Behind and science education: Opportunities, challenges and risks. *Elementary School Journal*, **106**, 476–477.

Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental Psychology*, **31**(5), 789–799.

Morrow, L. M., Pressley, M., Smith, J. K., & Smith, M. (1997). The effect of a literacy-based program integrated into literacy and science instruction with children from diverse backgrounds. *Reading Research Quarterly*, **32**(1), 54–76.

National Academy of Science. (2006). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: National Academy of Sciences, National Academy of Engineering and Institute of Medicine.

National Assessment of Educational Progress. (2007). *National Assessment of Educational Progress: The nation's report card*. Retrieved from http://nces.ed.gov/nationsreportcard/

National Institute of Child Health and Human Development. (2000). *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*. Washington, DC: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Child Health and Human Development.

National Institute of Child Health and Human Development, Early Child Care Research Network. (2004). Multiple pathways to early academic achievement. *Harvard Educational Review*, **74**(1), 1–29.

Palincsar, A. S., Collins, K. M., Marano, N. L., & Magnusson, S. J. (2000). Investigating the engagement and learning of students with learning disabilities in guided inquiry science teaching. *Speech, Language, and Hearing Services in Schools*, **31**(3), 240–251.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science*, **328**, 459–463.

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

Romance, N. R., & Vitale, M. R. (2001). Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications. *International Journal of Science Education*, **23**(4), 373–404.

Schatschneider, C., Buck, J., Torgesen, J. K., Wagner, R., Hassler, L., Hecht, S. A., & Powell-Smith, K. (2004). *A multivariate study of individual differences in performance on the reading portion of the Florida Comprehensive Assessment Test: A brief report*. Retrieved August 4, 2006, from http://www.fcrr.org/TechnicalReports/Multi_variate_study_december2004.pdf

Snow, C. E. (2001). *Reading for understanding*. Santa Monica, CA: RAND Education and the Science and Technology Policy Institute.

Snow, C. E. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Arlington, VA: RAND.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale* (4th ed.). Chicago: Riverside.

Torgesen, J. K. (2000). Individual differences in response to early intervention in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice*, **15**, 55–64.

Tuyay, S., Jennings, L., & Dixon, C. (1995). Classroom discourse and opportunities to learn: An ethnographic study of knowledge construction in a bilingual third-grade classroom. *Discourse Processes*, **19**, 75–110.

Van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, **328**, 453–456.

Von Secker, C. (2002). Effects of inquiry-based teacher practices on science excellence and equity. *Journal of Educational Research*, **95**(3), 151–160.

Williams, J. P., Hall, K. M., Lauer, K. D., Stafford, B., DeSisto, L. A., & deCani, J. (2005). Expository text comprehension in the primary grade classroom. *Journal of Educational Psychology*, **97**(4), 538–550.

Yore, L. D., Bisanz, G. L., & Hand, B. M. (2003). Examining the literacy components of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, **25**(6), 689–725.