

A RANDOMIZED CONTROL TRIAL OF SHARED EVALUATION PEDAGOGY: THE NEAR-TERM AND LONG-TERM IMPACT OF DIALOGICALLY ORGANIZED READING INSTRUCTION

MAREN AUKERMAN*, PAOLO C. MARTIN*, JOHN GARGANI** &
RICHARD D. MCCALLUM***

** Stanford University, ** Gargani + Company, *** University of California*

Abstract

This randomized control trial evaluated the impact of dialogically organized reading instruction provided twice per week over six months in small-group settings to fifth grade students, primarily English learners, assessed as needing additional support in comprehension development. Our dialogically organized instruction treatment, which we call Shared Evaluation Pedagogy (SHEP), was premised on: a) instigating dialogue about text, b) probing student thinking, and c) eschewing teacher evaluation. The 22 students in the SHEP treatment group demonstrated statistically significantly greater gains on near-term, researcher-administered measures of comprehension and decoding than did the 41 students in the control group. Treatment students also outperformed control students on two measures on the California Achievement Test 6 administered one year later—spelling (statistically significant) and reading (large enough to be practically significant, but not statistically significant). We close by arguing there are strong pedagogical reasons for shifting the prevailing reading comprehension intervention paradigm toward dialogically organized forms of instruction that invite students to think and reason with one another about texts.

Keywords: reading comprehension; dialogic instruction; reading intervention; English language learners

1. INTRODUCTION

While a number of researchers have documented the promise of dialogically organized instruction within literacy classrooms (Nystrand, 1997; Reznitskaya & Gregory, 2013), there are as yet few experimental or quasi-experimental studies to establish if such instruction yields transfer effects that children carry over into new literacy tasks (Wilkinson, Murphy, & Binici, 2015). Moreover, no such research to date has considered whether dialogically organized instruction is fruitful for teaching students specifically identified as low-achieving readers in an instructional intervention setting.

In this paper, we report on a randomized trial we conducted to evaluate the impact of a dialogically organized reading intervention on the academic performance of fifth grade students, most of whom were designated as English learners. The academic achievement of students who participated in the reading intervention, which we call Shared Evaluation Pedagogy (SHEP), was compared with that of students in a randomly assigned control group. We considered five outcome measures: two reading assessments administered by researchers immediately following the SHEP intervention and three standardized measures (in reading, spelling, and language) administered by the state one year later. We organized our study around our primary research question: was the SHEP intervention able to positively influence student achievement as reflected by these five measures? Our purpose in posing the question was to estimate the impact of the SHEP intervention specifically and to establish whether dialogically organized instruction might generally be a viable alternative to more explicit, direct forms of instruction that characterize most current comprehension interventions (e.g., Brown, Morris, & Fields, 2005; Johnson-Glenberg, 2000; Lubliner, 2004; Mason, 2004).

1.1 *The relationship between talk about text and student learning*

Nystrand (1997) proposes that “the underlying epistemology of classroom interaction defines the bottom line for learning: What ultimately counts is *the extent to which instruction requires students to think, not just report on someone else’s thinking*” (p. 72, emphasis in original). His correlational study of classroom discourse in eighth- and ninth-grade English classes found that classrooms that were organized for more substantive conversation resulted in more student learning of course content. A follow-up study of 64 middle and high school classrooms conducted with several other colleagues produced similar findings (Applebee, Langer, Nystrand, & Gamoran, 2003).

In keeping with Nystrand (1997), we highlight several distinctive features associated with dialogically organized instruction as we conceptualize it:

- *Authenticity*. Participants talk together in order to explore their own and each other’s textual perspectives. Students discuss ideas because they care about

them and have a stake in working them through. They have substantial authority to interpret texts for themselves; teachers seek to understand and make space for those interpretations, in part by posing *authentic questions*, to which they do not have a set answer in mind (Nystrand, 1997). When students pose questions, these are usually authentic as well, reflecting their textual curiosity (Aukerman, 2015).

- *Contingency*. Participants respond to one another in ways that unfold turn by turn: neither the forms of participation nor the topics to be discussed are pre-scripted (Boyd & Rubin, 2006). Teachers often engage in *uptake*, a kind of question that builds on a student's previous comment, in order to elicit elaboration and further reflection on a point a student has made (Nystrand, 1997).
- *Organic student-driven dialogue in which divergent textual perspectives come into conversation – collide, even*. Differing perspectives about the text's possible meanings are actively solicited by the teacher. Students talk about their ideas for extended stretches of time rather than shifting topics frequently or moving through a pre-set protocol of activities. They engage directly with each other's thinking in the talk that takes place, so the dialogue is characterized by *open discussion*, where students (or at least three classroom participants) engage in an exchange of ideas (Nystrand, 1997).

It is important to note that simply involving students in talk about text, as is done in a number of instructional approaches to comprehension, does not necessarily mean that the talk will be dialogically organized (O'Connor & Michaels, 2007). A number of approaches, particularly those that emphasize the explicit teaching of strategies to support student comprehension of text, share some features with dialogically organized instruction, but do not depend on the satisfaction of the above criteria. For example, reciprocal teaching (Palincsar & Brown, 1986; Takala, 2006) and inference training (McGee & Johnson, 2003), which do at times involve students talking directly with each other, may emphasize student-generated test questions (not fully satisfying the authenticity criterion), place no special emphasis on teacher uptake of student ideas (not fully satisfying the contingency criterion), and/or work from structured, pre-scripted protocols that deliberately shape what will be talked about when (not fully satisfying the organic student-driven dialogue criterion). As Wilkinson and Son (2011) have argued in their review of the historical turn in recent years toward dialogism in contemporary reading instruction, authentic dialogue no doubt is *possible* within some forms of strategy-based instruction (and, indeed, could potentially account for assessed comprehension gains more than student application of the strategy taught); still, they categorize such programs as conceptual precursors of dialogic teaching, not as part of the current wave of dialogic teaching, explaining what is distinctive in the current wave of dialogic teaching in these terms:

What is key to these more dialogic approaches is the juxtaposition of relative perspectives or discourses that gives rise to tension and sometimes conflict among different

voices. From a dialogic perspective, it is from the interaction and struggle among different, even competing, voices that meaning and understanding emerge.

Wilkinson and Son (2011) have argued that the move toward dialogically organized instruction is motivated by concerns about strategy instruction becoming so “mechanical” as to “inhibit generative learning” (p. 366) as well as being hard for teachers to learn and practice. Because we share these concerns and have insufficient space to provide a full review of all reading comprehension programs that involve student talk, we focus our subsequent literature review on discussion-based approaches anchored in the aforementioned criteria.

While there has been more research on discussion with students at the middle and high school levels (Nystrand, 2006), there are indications that discussion-based approaches may make a difference for children in elementary school as well. In socioeconomically disadvantaged primary-grade classrooms where there is more discussion and where teachers ask higher-level questions during discussions, students achieve at higher levels in their reading (Taylor, 2000); a similar pattern has been found for upper elementary-aged students (Bitter, 2009; McElhone, 2012). Kong & Fitch (2002/2003) found substantial gains in student performance on a range of comprehension-related measures over the course of a year in a classroom that based its reading instruction around student-led Book Club discussions (McMahon, 1997).

However, few experimental and quasi-experimental studies have examined the efficacy of discussion. Indeed, in Wilkinson, Murphy, and Binici’s 2015 review of the most extensively researched discussion approaches, they identified only nine experimental or quasi-experimental studies that considered whether text discussion had any transfer effect on literacy outcomes (i.e., outcomes with texts/materials that were not taught or otherwise introduced during the provided instruction). Among these studies, there were several (e.g., McKeown, Beck, & Blake, 2009) that reported significant positive outcomes; however, other studies (e.g., Reznitskaya et al., 2012) found no change in children’s assessed reading comprehension as a result of participation in discussion. Research with English learners has been more limited still, though one quasi-experimental study found that opportunities for discussion about a text were more effective than pre-modified written input at developing children’s assessed comprehension of that passage (Van den Branden, 2000). Thus, while there are reasons to be cautiously optimistic about the potential of dialogically organized discussions, further research is needed, particularly with low-achieving readers and English learners.

1.2 Intervention studies targeting assessed comprehension

The design of most pull-out literacy interventions appears sharply influenced by the premise that explicit instruction in reading skills and strategies is needed in order to improve assessed comprehension. Stahl (2004), for example, argues that “low achievers may need more explicit instruction” (p. 600) in comprehension strategies

than their higher-achieving peers (see also Mastropieri & Scruggs, 1997; Pressley & McCormick, 1995; Wilder & Williams, 2001). And explicit instruction, as operationalized in intervention programs, does not generally include an emphasis on dialogic discussion: in our search of the literature on reading interventions over the past twenty years, studies in which assessed comprehension for low-achieving readers was addressed through discussion-based approaches that met our criteria were notably absent, with the exception of Commeyras' (1992) exploratory study of seven students designated as learning disabled in a discussion-based small group.

A majority of pullout reading interventions do not target assessed comprehension directly; rather, they focus on phonics and word identification (Baker et al., 2000; Blanchman et al., 2004; Broaddus & Bloodgood, 1999; Craig, 2006; Compton et al., 2005; Fitzgerald, 2001; Hedrick & Pearish, 1999; MacKenzie, 2001; Morris et al., 2000; Santa & Høien, 1999; Schwartz, 2006; Torgeson et al., 1999; Vadasy, Sanders, & Peyton 2006) or fluency (Kuhn, 2004). Other studies (e.g., Jordan et al., 2000) have included a focus on vocabulary instruction as a lever for improving students' assessed comprehension. Of those studies that target global comprehension directly, the emphasis is generally on explicit teaching of comprehension strategies such as summarization, question generation and/or specific kinds of comprehension-related tasks, such as locating the main idea of a passage or making inferences about textual meaning (Brown et al., 2005; Leslie & Allen, 1999; O'Connor et al., 2002; Johnson-Glenberg, 2000; Lubliner, 2004; Mason, 2004; Meyer et al., 2002; McGee & Johnson, 2003). We were not able to locate intervention studies that provided opportunities for students to explore textual meaning through authentic, contingent, open dialogue.

Furthermore, it is notable that students considered to be less skilled as readers may have fewer opportunities for engaging in dialogically organized discussion even *within* their regular classroom settings (Applebee et al., 2003; Nystrand, 1997). Our study seeks to examine whether, when such instruction is provided in a pull-out intervention, students can benefit from it.

1.3 *Shared Evaluation Pedagogy*

As prelude to a description of our intervention pedagogy, we offer a brief vignette describing how one of the participating children, Maribel (a pseudonym), articulated the purpose of the small group discussions in which she was participating. Another child, a friend of hers, was not at the first few sessions, and the discussion facilitator asked the students to explain to the newcomer what they did during these reading discussions. Maribel had a simple way of explaining it. "We figure things out," she told her friend.

We could not have better articulated our own goals in the intervention. We sought to have the students collaboratively elaborate and explore textual hypothe-

ses, read on in search of evidence related to these hypotheses and revise them as they learned more in order to “figure things out” about what they were reading.

Each SHEP intervention session was focused on one piece of text, usually a short story or a picturebook that was read aloud by the students and/or the discussion facilitator during the session; longer texts might be completed over two sessions. Before the text was read, the facilitator provided a “book orientation” in which s/he briefly presented one or two key concepts or vocabulary words likely to spark interest and/or be unfamiliar to the students; there was minimal introduction to the book otherwise. The rest of the session involved reading the text, with the facilitator stopping for discussion after every several paragraphs. At each stop point, the teacher would either a) pause and wait for a student to comment or pose a question about what was read; b) pose an authentic question (Nystrand, 1997) about the story; or c) return to a previously discussed student idea about the text and ask students to think about that point in light of the new information that had just been read. Once a student comment had been elicited at a stop point, the facilitator would follow up by either asking for elaboration on that student’s comment or by asking others to respond – unless, of course, they spontaneously did so. The discussions typically ended with the teacher and students mentioning what had been most interesting to them about the text, or identifying something they would like to keep considering about the text.

The emphasis on having students create, extend and revise hypotheses in Shared Evaluation Pedagogy (SHEP) was maintained through adherence to three primary instructional principles:

Instigate. In keeping with the findings of Applebee and his colleagues, who found that open discussion took place most frequently when there were clear differences of opinion about textual meaning (Applebee et al., 2003), the texts selected for SHEP were high-interest texts that lent themselves to multiple interpretations. For example, groups read picturebooks such as *The Sweetest Fig* (Van Allsburg, 1993), in which a man turns into the dog he has been maltreating, or short stories such as Babbitt’s (1984) “The Imp in the Basket,” where both the characters in the story and the readers of the story wrangle with the question of whether a baby imp left on a clergyman’s doorstep was left by God or by the devil. Students were encouraged to articulate different interpretations, and there was no premium put on eventually coming to consensus about “the” meaning of the texts.

Probe by following, not leading. The discussion facilitator avoided the traditional Initiation-Response-Evaluation (I-R-E) discourse sequence, where the teacher poses questions with known answers, the students respond and the teacher evaluates those responses, typically moving on immediately to a new question and new topic (Mehan, 1982). Instead, the emphasis was on following students’ existing lines of thought, both by posing authentic questions (Nystrand, 1997) and by having students respond to each other’s ideas. Facilitator responses to students were highly contingent on student contributions; facilitators used uptake (Collins, 1986, Nystrand, 1997) as one strategy for maintaining thematic continuity and allowing

for elaboration of children's ideas. Facilitators deliberately minimized explicit instruction; aside from the book orientation, they clarified vocabulary primarily when students expressed curiosity about words during the readaloud and/or discussion. They also did not provide sentence frames (Fisher, 2008) or any similar kinds of scaffolding for the talk. Students had substantial freedom to talk about those aspects of the text they most cared to discuss.

Do not evaluate. Lastly, the facilitator did not provide evaluation that suggested that a student's hypothesis was right or wrong, even if the child's hypothesis suggested a potential "misreading" of the text. Less plausible hypotheses were handled in much the same way as more plausible hypotheses: the student was encouraged to explain, elaborate and justify based on textual evidence and prior knowledge; and other students were invited to respond. As Kelly (2007) has found, frequent evaluation of student ideas – particularly for low achieving readers – can decrease student engagement. The standard understanding of any one particular text was not the overriding goal, and it was understood that there would be times when student understandings would remain incomplete or non-standard. The aim was to develop facility with a process, not a pre-established understanding of the meaning of the text under discussion.

Appendix 1 provides a sample completed SHEP lesson plan, with the accompanying reflection from the discussion facilitator; Appendix 2 offers a transcript excerpt of a typical SHEP discussion. (A more comprehensive description of a SHEP dialogue can be found here: Aukerman, 2007).

1.4 The SHEP Intervention

Treatment students met in dialogically organized SHEP text discussion groups for two 50-minute sessions each week. Each group consisted of four to six students whose diagnostic profiles indicated grade level or near grade level word recognition skills, but whose comprehension scores as measured on the QRI-II (Leslie & Caldwell, 1995) fell from one to five years below grade level. Such a profile of reading (strong decoding skills, assessed comprehension not as strong) appears frequently among upper elementary students, particularly among emergent bilingual students reading in their emerging second language (Burgoyne, Kelly, Whiteley, & Spooner, 2009; Cain & Oakhill, 1999; Catts, Compton, Tomblin, & Bridges, 2012).

Each discussion group was designed to be heterogeneous, with students representing a spectrum of proficiencies in their assessed comprehension achievement. Groups met an average of roughly 40 times over the course of the year, in most cases replacing regular instruction during the school day. (We accommodated a few groups to be served after school at the request of the classroom teacher; as we explain later, we applied a statistical model enabling us to determine whether the additional instructional time those students received mattered in terms of student performance.)

The discussion groups were facilitated by graduate students enrolled in a master's program that prepared teachers to be reading specialists. All of them had studied dialogically organized instruction as part of their graduate program. (One was a recent graduate of that program who was pursuing doctoral work.) In addition, discussion facilitators received initial preparation and monthly training on the principles of Shared Evaluation Pedagogy (see above); preparation included joint observation and discussion of videotaped SHEP lessons.

We did not examine instructional discourse in the SHEP groups in terms of fidelity of implementation for both practical and theoretical reasons (cf. Buxton et al., 2015). On a practical level, our permission to work with students at the site did not include permission to collect audio or video data for research purposes. On a theoretical level, we agree with a number of researchers who have argued that dialogic teaching is about overall context and instructional stance; therefore, an analysis that only tracks discourse moves is likely to be an incomplete picture of the extent to which the discussion is dialogically organized (Boyd & Markarian, 2015; O'Connor & Michaels, 2007). Because we worked closely with the facilitators and informally observed their instruction throughout the year of the study, we are confident that they had a strong familiarity with SHEP principles, a commitment to a dialogic teaching stance, and the necessary skill to enact dialogically organized instruction on a consistent basis.

1.5 How might SHEP make a difference in terms of assessed comprehension?

It is worth noting that scholarly research around dialogic instruction is itself multi-voiced, with a number of different traditions and beliefs guiding what earns the right to be called dialogic (Matusov & Wegerif, 2014; Wegerif, 2008). For this reason, we in no way see our instantiation of dialogically organized instruction as the only one possible, and acknowledge that other scholars who choose to characterize a pedagogy as dialogic may make different choices around pedagogical principles and materials. Central to our own conceptualization of dialogic pedagogy, however, is that it is characterized by a stance rather than by a rigid sequence of moves (Boyd & Markarian, 2015; O'Connor & Michaels, 2007); this stance is one that honors student ideas about text and puts them in conversation with one another. We believe that such a stance may be ethically important in the classroom regardless of transferable literacy outcomes (Aukerman, 2013), but we also believe that honoring student textual ideas and putting them in dialogue potentially can be a meaningful lever for improving children's assessed comprehension.

Honoring student ideas is doubtless easier to do when students' textual understandings are identifiably text-based or otherwise clearly defensible. Many discussion-based approaches are uncomfortable with letting students' nonstandard ideas remain, particularly at the end of a discussion. In Van den Branden's (2000) study of multilingual fifth graders, which looked at the effects on reading comprehension of "collective meaning negotiation" (an approach that sought to let students take

more of their own initiative in figuring out textual meaning), the teacher-researcher “did give hints if necessary and endorsed the correct meaning of difficult input if it had been formulated by a pupil” (p. 433). Another teacher-researcher told her class in the context of their Book Club discussions, ““The outlandish ideas need to stay out of the discussion”” (Kong & Fitch, 2002/2003, p. 358). And, while some scholars of dialogic pedagogy propose that “the dialogic process guards against errors in substantive conclusions, as the group continually self-corrects” (Reznitskaya & Gregory, 2013, p. 117), we have not found that honoring student ideas about text will necessarily lead to consensus, a single textual understanding, or the correction of potentially inaccurate conclusions (Aukerman, forthcoming). Yet we maintain that children’s alternate, non-standard understandings are reflective of active textual sensemaking, also a kind of meaningful comprehension (Aukerman, 2013). For this reason, we refer in this paper to student performance on reading comprehension measures that allow only one correct response as “assessed comprehension.”

We theorized that dialogue about student textual ideas – their *actual* understandings of texts no matter how apparently imperfect or even “outlandish” – would support students’ long-term, transferable abilities to perform well in *assessed* comprehension. We believe that learning to repair hypotheses involves investigating them throughout the entire text being read, depending on the text (and not a teacher) to evaluate the plausibility of the hypothesis. Children who voice non-standard understandings in the context of discussion can facilitate that kind of active evaluation on the part of their peers. We believe that steering the group to a more standard interpretation might produce a short-term bump in assessed comprehension of a particular text under discussion, but it often deprives the group of the opportunity to engage in intellectual work that builds student capacity to make evidence-based decisions about what kinds of textual interpretations are most plausible. Thus, we see making space for children’s non-standard textual understandings to be explored during classroom talk as not only compatible with a long-term goal of improving assessed comprehension, but actually facilitative of it.

1.6 How might SHEP make a difference in terms of additional literacy outcomes?

In addition to theorizing that children’s assessed reading comprehension would improve through participation in a SHEP discussion group, we also believed that two additional literacy skills might be affected by the SHEP intervention: decoding and spelling. We chose to look at decoding because, while the program did not involve explicit teaching of decoding skills, students did do some reading aloud during SHEP discussions. Moreover, there is considerable evidence that decoding and assessed comprehension are demonstrably correlated in children. Although there has been more work documenting the dependence of assessed comprehension on decoding ability (e.g., Shankweiler et al., 1999) than vice versa, it is also plausible that, as children’s reading comprehension increases, they both read more and en-

gage with more varied texts, which in turn can promote decoding skill (cf. Stanovich, 1986).

While spelling achievement might initially appear less related to the SHEP intervention, recent research on reciprocal effects of assessed reading comprehension and spelling indicate that assessed comprehension plays a substantial contributing role in students' spelling ability (Retelsdorf & Köller, 2014). Although the reasons for this contribution have not been empirically determined, Frith (1985) has proposed that older readers gain information about spelling from what they read; she has reasoned that the more children read and comprehend, the better they are able to spell. We thus believed that, if students were developing as comprehenders through their participation in SHEP dialogues, their spelling abilities were likely to improve as well.

2. METHOD

2.1 Site and Participants

The students in the study attended Sheridan Elementary School (a pseudonym), a public school situated in an urban district in California. Sheridan served close to 1,000 students during the study year, of whom 80% received free or reduced lunch and 76% were designated English Learners (ELs). Treatment and control students were drawn from six self-contained fifth-grade classrooms where reading instruction was provided by experienced, traditionally certified teachers with a structured basal text in English, *Open Court Reading* (Bereiter et al., 2002). Two classrooms were designated as transitional bilingual classrooms but provided reading instruction in English, as did the other four classrooms.

2.2 Design

We implemented a randomized control trial within a larger effort to develop student literacy. Our primary duty was to support students. At the same time, we set out to conduct a study that many quantitative researchers would consider more rigorous than previous studies on this topic. This required us to balance pedagogical rigor with experimental rigor, differentiated instruction with a fixed research protocol, and our duty to students with our commitment to a method. To help us achieve balance, we took advantage of a randomization procedure that is more complex than it might have been had we privileged research above other concerns.

We selected the study sample at the start of the school year using a four-step process, illustrated in Figure 1. First, trained graduate student researchers individually assessed all 179 fifth graders at Sheridan using word recognition, oral reading, and non-fiction silent reading tests drawn from the Diagnostic Assessment of Reading (DAR; Roswell, 2005) and the Qualitative Reading Inventory-II (QRI-II; Leslie & Caldwell, 1995). Second, we used the assessment results and information provided

by teachers to identify students who 1) spoke English well enough to understand the directions given to them by the graduate students and to complete the assessment, 2) scored one or more grade levels below their current grade level in reading comprehension, and 3) were not currently receiving supplemental special education services for reading. One hundred twenty-four students met these selection criteria.

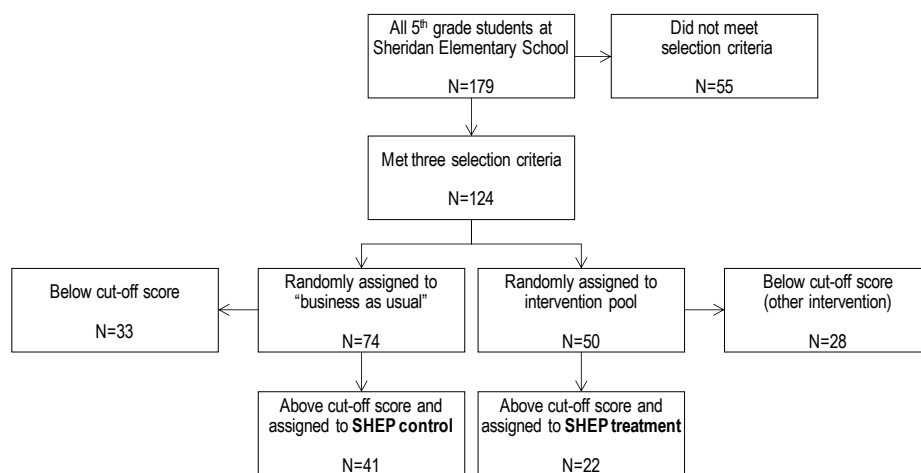


Figure 1. The Random Assignment of Students to SHEP Treatment and Control Groups

Third, we randomly assigned 50 of these students to receive a reading intervention of some kind (the intervention pool). We blocked on teacher when we randomized. Thus, 7 to 10 students from each of the 6 participating classrooms were assigned at random to the intervention pool, more or fewer depending on the total number of selected students in the classroom.

Fourth, we constructed a single scale from the word recognition, oral reading, and non-fiction reading comprehension scores. For each classroom, we picked a point on the scale that divided the intervention pool into two approximately even groups; the point on the scale varied by classroom. This resulted in a total of 28 students on the lower end of the scale who had weak word recognition skills. They were placed in a parallel intervention program where they received one-on-one tutoring that strongly emphasized word recognition in the context of guided reading. We do not discuss this intervention here, only note that it formed part of a larger effort to provide differentiated support to students and that it was accommodated by our research design. The remaining 22 students on the higher end of the scale had relatively robust word recognition skills but weak assessed comprehension skills. They were assigned to dialogically organized SHEP instruction, and

they formed the SHEP treatment group. Of the 74 students who met our three selection criteria but were not randomly assigned to the intervention pool (they experienced “business as usual”), 41 fell on the higher end of the scale using the same cut-off points. They formed the SHEP control group.

After one semester (3 months) of twice-weekly instruction, students in the SHEP program and in the parallel intervention program were reassessed on the same measures to determine what kind of further intervention, if any, was needed for them to be able to read at grade level. Six students who had been receiving SHEP instruction were exited from the program. We included them in the treatment group and did not adjust for their shorter intervention.

Table 1 provides additional demographic information on the study sample. Overall, the sample was predominantly Latino (79%) and English learners (86%). Across treatment and control groups, student characteristics were comparable, but not perfectly so. A higher proportion of the treatment students were identified by the schools as Latino and designated as English learners, and treatment students on average had slightly lower pre-intervention word recognition and reading comprehension scores. These differences suggest that treatment students may have been more difficult to instruct in reading, reducing treatment and control differences.

Table 1. Demographic Information on the Study Sample

Variable	Control		Treatment		All	
	<i>N</i>	<i>Percent</i>	<i>N</i>	<i>Percent</i>	<i>N</i>	<i>Percent</i>
Sample Size	41	65%	22	35%	63	100%
Gender						
Male	21	33%	10	16%	31	49%
Female	20	32%	12	19%	32	51%
Ethnicity						
Latino	28	44%	22	35%	50	79%
Asian/Pacific	7	11%	0	0%	7	11%
African American	2	3%	0	0%	2	3%
Other/Unknown	4	6%	0	0%	4	6%
English Learner Status						
EL	33	80%	21	95%	54	86%
Not EL	8	20%	1	5%	9	14%
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Pre-Intervention Scores						
Word Recognition	4.3	1.1	3.9	1.1	4.2	1.1
Oral Reading	5.7	0.8	5.7	0.6	5.7	0.7
Comprehension	3.1	1.0	2.5	1.2	2.9	1.1

2.3 Outcome Measures

We evaluated the impact of the program with five outcome measures. All reflected transfer effects of the intervention, as the texts used in the assessments differed from those used in the intervention. Two measures were constructed by rescaling the end-of-year word recognition, oral reading, and comprehension assessments delivered to all treatment and control students and mapping them to two dimensions—decoding and comprehension—using a two-dimensional item response model (Frank Rijmen & Briggs, 2004). The internal consistencies of these measures were 0.80 and 0.85, respectively, as estimated by the marginal maximum likelihood reliability (Kim & Wilson, 2009; Mislevy, Beaton, Kaplan, & Sheehan, 1992). This is analogous to Cronbach's alpha and estimated within the item response model using a proportion-of-variance approach.

The remaining three measures were standardized test scores. Although we did not expect SHEP to impact all dimensions of language development, we included all language-related measures from the California Achievement Tests (CAT/6) as outcome measures to help establish what might be possible. These measures are referred to as Reading, Language, and Spelling. They were administered 12 months *after* the intervention concluded, allowing us to gauge the program's sustained impact on measures with substantial validity and reliability evidence.

2.4 Statistical Model

Impact estimates for all five outcome measures were estimated simultaneously using a type of hierarchical linear model (HLM; Raudenbush & Bryk, 2002), a well-known class of model from which contemporary education researchers commonly draw. The multivariate hierarchical linear model we use (Tate and Pituch, 2007), while perhaps less common than some other statistical methods, was our preferred approach because it has the ability to impute missing test scores using a maximum likelihood approach. That is, it nests outcome measures within students by borrowing information from available test scores to impute those that are missing and incorporating the uncertainty of the imputed measures in the larger model. In the context of the hierarchical linear model that we used, the maximum likelihood approach typically yields results that are similar to or better than others (see Enders, 201, pp. 336-340). A total of 53 out of 315 outcome measures (17%) were missing either because students moved to other schools in the subsequent year or were unavailable for testing in the study year. No pre-intervention covariates were missing. Table 2 presents the pattern of missing and available outcome measures in our data.

Table 2. The Pattern of Missing (M) and Available (A) Outcome Measures

Condition	(N) Students	(N) Measures	CAT/6 Standardized Test			Constructed Measures	
			Reading	Language	Spelling	Decoding	Comprehension
Control	13	2	M	M	M	A	A
	1	3	A	A	A	M	M
	27	5	A	A	A	A	A
Treatment	4	2	M	M	M	A	A
	18	5	A	A	A	A	A

3. RESULTS

Three alternative models were fit that incorporated different student-level variables. Model 1 controlled for pre-intervention covariates, which were decoding and comprehension scores constructed from individual assessments conducted at the start of the study year. Model 2 additionally controlled for teachers (using dummy variables), allowing us to take blocking into account. Finally, Model 3 controlled for pre-intervention covariates and any additional instructional time SHEP treatment students received (i.e., students who received the intervention after school). This model allows us to gauge the extent to which outcome measures were affected by SHEP itself versus instructional time. Seven students received additional instruction because their teachers believed their interests were best served by using SHEP after school to augment in-class instruction. This was another real-world factor that we incorporated into our experimental design. Because there is high level of dependence between pre-intervention covariates, teacher, and additional time, all three variables could not be incorporated into a single model. We identified the best fitting model as the one that minimized the Akaike and Bayesian information criteria (AIC and BIC: see Burnham & Anderson, 2002, for a discussion of how information criteria may be used to evaluate model fit, which is analogous to how less reliable adjusted R-squared values are sometimes used to evaluate competing regression models).

Table 3 presents the impact estimates for these three models. The standardized effect sizes were computed by dividing impact estimates in their original units by the pooled standard deviation of the corresponding measure. The pooled standard deviation was taken to be the standard deviation of the residuals for each measure in a multivariate hierarchical linear model that controlled for condition only (for reading, language, spelling, decoding, and comprehension these values are 31.96, 31.72, 48.81, 1.29, and 5.26, respectively). The estimates for Models 1 and 2 are similar. Impact estimates for decoding, comprehension, and CAT/6 spelling are large, falling in the 0.76 to 0.94 SD range, and statistically significant at the .05 level

with and without the Benjamini–Hochberg (BH) adjustment for multiple inference (Benjamini & Hochberg, 1995). The impact estimates for CAT/6 Reading are smaller, falling in the range of 0.15 to 0.18, and although large enough to be practically significant, they are not statistically significant. The impact estimate for CAT/6 Language is effectively 0.

Table 3. Estimates of SHEP's Impact on Students as Reflected by Five Literacy Measures

						Stat. Sig at .05 Level		Model Selection	
Outcome Measure	St. ES	St. Error	T-Ratio	Appr. DF	p-Val.	No Adj.	BH Adj.	AIC	BIC
Model 1: Pre-Intervention Covariates (Decoding & Comprehension)									
CAT/6 Reading	0.181	0.320	0.565	59	0.574				
CAT/6 Language	-0.030	0.337	-0.088	59	0.931				
CAT/6 Spelling	0.812	0.243	3.348	59	0.002	*	*	1810	1863
Decoding	0.758	0.234	3.233	59	0.002	*	*		
Compre- hension	0.938	0.248	3.779	59	0.001	*	*		
Model 2: Pre-Intervention Covariates & Teacher (Dummy Variables)									
CAT/6 Reading	0.153	0.330	0.462	54	0.645				
CAT/6 Language	-0.014	0.321	-0.042	54	0.967				
CAT/6 Spelling	0.791	0.255	3.104	54	0.003	*	*	1717	1824
Decoding	0.800	0.247	3.239	54	0.002	*	*		
Compre- hension	0.897	0.262	3.420	54	0.002	*	*		

Model 3: Pre-Intervention Covariates & Additional Time (Dichotomous Variable)

CAT/6							
Reading	0.145	0.363	0.399	58	0.691		
CAT/6							
Language	-0.086	0.383	-0.225	58	0.823		
CAT/6						1788	1852
Spelling	0.748	0.275	2.718	58	0.009	*	*
Decoding	0.487	0.256	1.898	58	0.062		
Compre- hension	0.764	0.279	2.739	58	0.009	*	*

Note. No Adj. = Cutoff for statistical significance of all tests was .05, BH Adj. = Cutoff for statistical significance based on the Benjamini–Hochberg adjustment for multiple inference, AIC = Akaike information criterion for model selection, BIC = Bayesian information criterion for model selection.

The results for Models 2 and 3 are similar with the exception of decoding. When additional time is taken into account, the standardized effect size decreases to 0.49 SD. This is substantial on a practical level, but not large enough to trigger statistical significance. Given the similarity in the results for Models 2 and 3, additional time is not a plausible rival hypothesis that explains the differences between the treatment and control group.

Of the three models, Model 2 is preferred because it minimizes both the AIC and BIC criteria. Consequently, we use the impact estimates from this model to characterize the performance of the SHEP intervention. In the short term, SHEP improved the decoding and comprehension of students as reflected by statistically significant differences in the constructed measures. Over a 12-month period, SHEP also had a sustained impact on literacy ability as reflected by the CAT/6 test in spelling (statistically significant) and possibly reading (large enough to be practically significant but not statistically significant).

4. DISCUSSION

The results of our study indicate that students who participated in a dialogically organized SHEP intervention had transfer effects in their assessed ability to comprehend texts. Their significant achievement gains call into question what has become almost a truism in the world of reading comprehension instruction—that students who do not perform well in assessed comprehension need explicit, step-by-step guidance in order to make needed gains as readers (Pressley & McCormick, 1995; Stahl, 2004; Wilder & Williams, 2001). Our findings indicate that dialogically organized student-centered text discussion with minimal explicit instruction can

have significant positive effects as an intervention on comprehension and on decoding skills for students whose word identification skills are relatively robust but whose assessed comprehension is not as strong.

SHEP students did not outperform control students on the language portion of the CAT/6, but they did outperform control students on the CAT/6 in spelling at a statistically significant level nearly a year after the completion of the intervention. Given that our assessment did not explicitly teach spelling (or focus to any significant extent on writing), we believe these findings are most likely attributable to their stronger assessed comprehension; student performance fits with Frith's (1985) hypothesis that students whose assessed comprehension is stronger develop strong spelling skills in part through more exposure and attention to words they are reading, and Retelsdorf and Köller's (2014) finding that better assessed comprehension contributes to better spelling but not vice versa.

SHEP treatment students also trended toward better performance on the CAT/6 reading subtest. While these findings were not statistically significant, we believe the trends may be of practical significance, particularly when considered within the context of the strong performance of SHEP students on other measures. The CAT/6 reading test may not have been sufficiently sensitive to capture assessed comprehension differences between the populations, particularly given the relatively small sample size of the study.

There are at least two ways that SHEP pedagogy may have contributed to the encouraging literacy outcomes on the constructed measures and on the CAT/6. First, SHEP students were regularly encouraged to evaluate the textual claims of their peers, giving them both authentic reasons for reading closely and the opportunity to see how others were making sense of the text. Second, students in the small groups may have found opportunities to discuss text engaging, and may well have found the rich, ambiguous texts read in the small group engaging; we informally observed that many students relished coming to the discussions. Since some scholars (Nystrand & Gamoran, 1991) have proposed that higher substantive engagement contributes to greater substantive learning, we believe that active engagement could have been a lever for change here.

We note several limitations of our study. First, while the discussion facilitators in this study had considerable experience and preparation in teaching dialogically, we did not record their language and are thus unable to decisively determine what kinds of language patterns were most typical of their instruction. Future research should closely investigate observed language patterns in relationship to student outcomes. A second potential limitation is that students in the control group did not participate in a competing instructional intervention. However, our Model 3 findings indicate that students who received the intervention after school (additional instructional time) versus during the regular school day (without additional instructional time) performed similarly; for this reason, we do not believe that additional instructional time was the key factor determining growth. Moreover, other interventions with fifth graders that have involved as much as 90 hours of addi-

tional instruction have had lackluster effects on reading achievement, particularly in the realm of assessed comprehension (Torgeson et al., 2006). Those findings lead us to believe that the critical dimension contributing to student learning in this study was the dialogic nature of the instruction students received in the SHEP treatment group.

Finally, we note that our findings may not be applicable to intervention settings where teachers have less knowledge and preparation that would enable them to provide dialogically organized instruction. Given that dialogically organized instruction is not easy to learn for teachers (Caughlan et al., 2013), it is not clear what kind of support intervention teachers would need in order to provide effective dialogically organized teaching in a SHEP intervention context. Further research in teacher preparation and professional development contexts should explore this question.

Our study adds to the modest, but growing, body of evidence that some forms of dialogically organized instruction can have transfer effects on children's literacy abilities (Wilkinson et al., 2015). The findings provide an encouraging indication that such effects can be long-term (lasting at least a year beyond the end of the intervention), and that populations involving a high percentage of English learners can benefit. Furthermore, it indicates that lower-achieving readers, like their higher-achieving peers, are capable of learning from dialogically organized conversations about text. More research is needed to establish how such an intervention might perform vis-à-vis other specific kinds of instructional interventions, such as explicit strategy instruction (cf. McKeown et al., 2009). We also hope future research will examine whether children who receive a dialogically organized intervention that offers additional instructional time benefit more than students who receive the same intervention without additional instructional time (i.e., during the school day). Finally, we wonder whether simply changing the nature of regular reading instruction might matter for students with low assessed comprehension: could dialogically organized instruction in the regular classroom have similar benefits to a dialogically organized pull-out intervention model?

We close by noting that SHEP offers a number of important advantages that make it worth considering as a preferred form of instructional intervention for children, like those in this study, whose word recognition skills are relatively robust but whose assessed comprehension is not as strong. SHEP does not require homogeneous grouping, special curricular sequencing, or even a particular set of student or teacher materials; it can be undertaken with texts selected for their ability to interest, inform and challenge particular, academically heterogeneous groups of students rather than for their utility in teaching particular comprehension strategies. Given that dialogically organized instruction may foster students' sense of self-efficacy as readers across time more than monologically organized instruction, particularly for lower-achieving students (Aukerman & Chambers Schuldt, 2015), there may be affective as well as strictly academic reasons for organizing literacy intervention instruction dialogically.

Finally, the emphasis in SHEP is on learning to engage thoughtfully *with others* around text by responding to peer ideas as one develops one's own. It is not simply what one can do with text, but what one can do in conversation with others around text, that matters. We maintain that this may be at least as significant a long-term objective as the improvement of students' reading as assessed comprehension, and look forward to further research in the field that investigates such outcomes.

REFERENCES

- Anderson, E. B. (1995). Polytomous Rasch models and their estimation. In G. H. Fischer, & I. W. Molenaar (Eds.), *Rasch Models: Foundations, recent developments and applications* (pp. 271-292). New York: Springer. http://dx.doi.org/10.1007/978-1-4612-4230-7_15
- Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, 40(3), 685-730. <http://dx.doi.org/10.3102/00028312040003685>
- Aukerman, M. (2007). When reading it wrong is getting it right: Shared Evaluation Pedagogy among "struggling" fifth grade readers. *Research in the Teaching of English*, 42(1), 56-103.
- Aukerman, M. (2013). Rereading comprehension pedagogies: Toward a dialogic teaching ethic that honors student sensemaking. *Dialogic Pedagogy: An International Online Journal*, 1(1). <http://dx.doi.org/10.5195/dpj.2013.9>
- Aukerman, M. (2015). How should readers develop across time? Mapping change without a deficit perspective. *Language Arts*, 93(1), 57-64.
- Aukerman, M., & Chambers Schuldt, L. (2015). Children's perceptions of their reading ability and epistemic roles in monologically and dialogically organized bilingual classrooms. *Journal of Literacy Research*, 47(1), 115-145. <http://dx.doi.org/10.1177/1086296x15586959>
- Aukerman, M., & Chambers Schuldt, L. (forthcoming). Closely reading "reading closely". *Language Arts*, 93(4).
- Babbitt, N. (1984). The imp in the basket. In *Junior Great Books Series 4* (Vol. 2). Chicago, IL: Great Books Foundation.
- Baker, S., Gersten, R., & Keating, T. (2000). When less may be more: A 2-year longitudinal evaluation of a volunteer tutoring program requiring minimal training. *Reading Research Quarterly*, 35(4), 494-519. <http://dx.doi.org/10.1598/RRQ.35.4.3>
- Baumann, J. F., Edwards, E. C., Font, G., Tereshinski, C. A., Kame'enui, E. J., & Olejnik, S. (2002). Teaching morphemic and contextual analysis to fifth-grade students. *Reading Research Quarterly*, 37(2), 150-176. <http://dx.doi.org/10.1598/RRQ.37.2.3>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289-300. <http://dx.doi.org/10.2307/2346101>
- Bereiter, C., Brown, A., Campione, J., Carruthers, I., Case, R., Hirshberg, J., Adams, M.J., McKeough, A., Pressley, M., Roit, M., Scardamalia, M., Stein, M., & Treadway, G.H. (2002). *Opencourt reading*. Columbus, OH: SRAMcGraw-Hill.
- Bitter, C., O'Day, J., Gubbins, P., & Socias, M. (2009). What works to improve student literacy achievement? An examination of instructional practices in a balanced literacy approach. *Journal of Education for Students Placed at Risk*, 14, 17-44.
- Boyd, M., & Markarian, W. (2015). Dialogic teaching and dialogic stance: Moving beyond interactional form. *Research in the Teaching of English*, 49(3), 272-296.
- Boyd, M., & Rubin, D. (2006). How contingent questioning promotes extended student talk: a function of display questions. *Journal of Literacy Research*, 38(2), 141-169. http://dx.doi.org/10.1207/s15548430jlr3802_2

- Broaddus, K., & Bloodgood, J. W. (1999). "We're supposed to already know how to teach reading": Teacher change to support struggling readers. *Reading Research Quarterly*, 34(4), 426-451. <http://dx.doi.org/10.1598/RRQ.34.4.3>
- Brown, K., Morris, D., & Fields, M. (2005). Intervention after grade 1: Serving increased numbers of struggling readers effectively. *Journal of Literacy Research*, 37(1), 61-94. http://dx.doi.org/10.1207/s15548430jlr3701_3
- Burgoyne, K., Kelly, J. M., Whiteley, H. E., & Spooner, A. (2009). The comprehension skills of children learning English as an additional language. *British Journal of Educational Psychology*, 79(4), 735-747. <http://dx.doi.org/10.1348/000709909X422530>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.
- Buxton, C. A., Allestaht-Snyder, M., Kayumova, S., Aghasaleh, R., Choi, Y.-J., & Cohen, A. (2015). Teacher agency and professional learning: Rethinking fidelity of implementation as multiplicities of enactment. *Journal of Research in Science Teaching*, 52(4), 489-502. <http://dx.doi.org/10.1002/tea.21223>
- Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, 11(5-6), 489-503. <http://dx.doi.org/10.1023/A:1008084120205>
- Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, 104(1), 166-181. <http://dx.doi.org/10.1037/a0025323>
- Caughlan, S., Juzwik, M. M., Borsheim-Black, C., Kelly, S., & Fine, J. G. (2013). English teacher candidates developing dialogically organized instructional practices. *Research in the Teaching of English*, 47(3), 212-246.
- Chinn, C. A., Anderson, R. C., & Waggoner, M. A. (2001). Patterns of discourse in two kinds of literature discussion. *Reading Research Quarterly*, 36(4), 378-411. <http://dx.doi.org/10.1598/RRQ.36.4.3>
- Collins, J. (1986). Differential instruction in reading groups. In J. Cook-Gumperz (Ed.), *The social construction of literacy* (pp. 117-137). Cambridge: Cambridge University Press.
- Commeyras, M. (1992). *Dialogical-thinking reading lessons: Promoting critical thinking among "learning-disabled" students*. Technical report no. 553. Urbana: Illinois University, Center for the Study of Reading.
- Compton, D., Olinghouse, N., Elleman, A., Vining, J., Appleton, A., Vail, J., et al. (2005). Putting transfer back on trial: Modeling individual differences in the transfer of decoding-skill gains to other aspects of reading acquisition. *Journal of Educational Psychology*, 97(1), 55-69. <http://dx.doi.org/10.1037/0022-0663.97.1.55>
- Craig, S. A. (2006). The effects of an adapted interactive writing intervention in kindergarten children's phonological awareness, spelling, early reading development: A contextualized approach to instruction. *Journal of Educational Psychology*, 98(4), 714-731. <http://dx.doi.org/10.1037/0022-0663.98.4.714>
- Eeds, M., & Wells, D. (1989). Grand conversations: An exploration of meaning construction in literature study groups. *Research in the Teaching of English*, 23(10), 4-29.
- El-Dinary, P., & Schuder, T. (1993). Seven teachers' acceptance of transactional strategies instruction during their first year using it. *The Elementary School Journal*, 94(2), 207-219.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Publications.
- Fisher, D., Frey, N., & Rothenberg, C. (2008). *Content-area conversations: how to plan discussion-based lessons for diverse language learners*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Fitzgerald, J. (2001). Can minimally trained college student volunteers help young at-risk children to read better? *Reading Research Quarterly*, 36(1), 28-47. <http://dx.doi.org/10.1598/RRQ.36.1.2>
- Frank Rijmen, & Briggs, D. (2004). Multiple person dimensions and latent item predictors. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: a generalized linear and nonlinear approach* (pp. 247-259). New York: Springer.
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In J. M. K. Patterson, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301-330). London, UK: Lawrence Erlbaum Associates.

- Hammerberg, D. D. (2004). Comprehension instruction for socioculturally diverse classrooms: A review of what we know. *The Reading Teacher*, 57(7), 648-658.
- Hedrick, W. B., & Pearish, A. B. (1999). Good reading instruction is more important than who provides the instruction or where it takes place. *The Reading Teacher*, 52(7), 716-727.
- Jewell, T. A., & Pratt, D. (1999). Literature discussions in the primary grades: Children's thoughtful discourse about books and what teachers can do to make it happen. *The Reading Teacher*, 52(8), 842-850.
- Johnson-Glenberg, M. (2000). Training reading comprehension in adequate decoders/poor comprehenders: Verbal versus visual strategies. *Journal of Educational Psychology*, 92(4), 772-782. <http://dx.doi.org/10.1037/0022-0663.92.4.772>
- Jordan, G. E., Snow, C. E., & Porche, M. V. (2000). Project ease: The effect of a family literacy project on kindergarten students' early literacy skills. *Reading Research Quarterly*, 35(4), 524-546. <http://dx.doi.org/10.1598/RRQ.35.4.5>
- Kelly, S. (2007). Classroom discourse and the distribution of student engagement. *Social Psychology of Education*, 10, 331-352. <http://dx.doi.org/10.1007/s11218-007-9024-0>
- Kim, S. C., & Wilson, M. (2009). A comparative analysis of the ratings in performance assessment using generalizability theory and the many-facet Rasch model. *Journal of Applied Measurement*, 10(4), 408-423.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.
- Kong, A., & Fitch, E. (2002/2003). Using book club to engage culturally and linguistically diverse learners in reading, writing and talking about books. *The Reading Teacher*, 56(4), 352-362.
- Kuhn, M. (2004). Helping students become accurate, expressive readers: Fluency instruction for small groups. *The Reading Teacher*, 58(4), 338-344. <http://dx.doi.org/10.1598/RT.58.4.3>
- Langer, J. (1994). *A response-based approach to reading literature*. Albany, NY: National Research Center on Literature Teaching and Learning.
- Leslie, L., & Allen, L. (1999). Factors that predict success in an early literacy intervention project. *Reading Research Quarterly*, 34(4), 404-424. <http://dx.doi.org/10.1598/RRQ.34.4.2>
- Leslie, L., & Caldwell, J. (1995). *Qualitative reading inventory - II*. New York: HarperCollins.
- Lewis, C. (1993). "Give people a chance": Acknowledging social differences in reading. *Language Arts*, 70, 454-461.
- Lubliner, S. (2004). Help for struggling readers. *The Reading Teacher*, 57(5), 430-438.
- MacKenzie, K. (2001). Using Literacy Booster Groups to maintain and extend Reading Recovery success in the primary grades. *The Reading Teacher*, 55(3), 222-234.
- Mason, L. H. (2004). Explicit self-regulated strategy development versus reciprocal questioning: Effects on expository reading comprehension among struggling readers. *Journal of Educational Psychology*, 96(2), 283-296. <http://dx.doi.org/10.1037/0022-0663.96.2.283>
- Mastropieri, M., A., & Scruggs, T. E. (1997). Best practices in promoting reading comprehension in students with learning disabilities. *Remedial & Special Education*, 18(4), 197-216. <http://dx.doi.org/10.1177/074193259701800402>
- Matusov, E., & Wegerif, R. (2014). Dialogue on 'Dialogic Education': Has Rupert gone over to 'the Dark Side'? *Dialogic Pedagogy: An International Online Journal*, 2, E1-E20. <http://dx.doi.org/10.5195/dpj.2014.78>
- McElhone, D. (2012). Tell us more: Reading comprehension, engagement, and conceptual press discourse. *Reading Psychology*, 33(6), 525-561. <http://dx.doi.org/10.1080/02702711.2011.561655>
- McGee, A., & Johnson, H. (2003). The effect of inference training on skilled and less skilled comprehenders. *Educational Psychology*, 23(1), 49-59. <http://dx.doi.org/10.1080/0144341022000022924>
- McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, 44(3), 218-253. <http://dx.doi.org/10.1598/RRQ.44.3.1>
- McMahon, S. I. (1997). Book clubs: Contexts for students to lead their own discussions. In S. I. McMahon, & T. E. Raphael (Eds.), *The Book Club connection: Literacy learning and classroom talk* (pp. 89-106). New York: Teachers College Press.

- Mehan, H. (1982). The structure of classroom events and their consequences for student performance. In P. Gilmore, & A. A. Glatthorn (Eds.), *Children in and out of school: Ethnography and education*. Washington, DC: Center for Applied Linguistics.
- Meyer, B., Theodourou, E., Brezinski, K., Middlemiss, W., McDougall, J., & Bartlett, B. (2002). Effects of structure strategy instruction delivered to fifth-grade children using the internet with and without the aid of older adult tutors. *Journal of Educational Psychology*, 94(3), 486-519. <http://dx.doi.org/10.1037/0022-0663.94.3.486>
- Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133-161. <http://dx.doi.org/10.1111/j.1745-3984.1992.tb00371.x>
- Moore, A. (1996). Assessing young readers: Questions of culture and ability. *Language Arts*, 73, 306-316.
- Morris, D., Tyner, B., & Perney, J. (2000). Early Steps: Replicating the effects of a first-grade reading intervention program. *Journal of Educational Psychology*, 92(4), 681-693. <http://dx.doi.org/10.1037/0022-0663.92.4.681>
- Murphy, P., & Edwards, M. (2005, April). *What the studies tell us: A meta-analysis of discussion approaches*. Paper presented at the Annual meeting of the American Educational Research Association, Montreal, Canada.
- Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.
- Nystrand, M. (2006). Research on the role of classroom discourse as it affects reading comprehension. *Research in the Teaching of English*, 40(4), 392-412.
- Nystrand, M., & Gamoran, A. (1991). Instructional discourse, student engagement and literature achievement. *Research in the Teaching of English*, 25(3), 261-290.
- O'Connor, C., & Michaels, S. (2007). When is dialogue dialogic? *Human Development*, 50(5), 275-285. <http://dx.doi.org/10.1159/000106415>
- O'Connor, R. E., Bell, K. M., Harty, K. R., Larkin, L. K., Sackor, S. M., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology*, 94(3), 474-485. <http://dx.doi.org/10.1037/0022-0663.94.3.474>
- Palincsar, A. S., & Brown, A. L. (1986). Interactive teaching to promote independent learning from text. *The Reading Teacher*, 39, 771-777.
- Pinkney, J. (2000). *Aesop's fables*. New York: SeaStar Books.
- Pressley, M., & El-Dinary, P. (1997). What we know about translating comprehension-strategies instruction research into practice. *Journal of Learning Disabilities*, 30(5), 486-489, 512. <http://dx.doi.org/10.1177/002221949703000504>
- Pressley, M., & McCormick, C. (1995). *Advanced educational psychology*. New York: Harcourt Brace.
- Rabinowitz, P. J. (1998). "A thousand times and never like": Re-reading for class. In P. J. Rabinowitz & M. W. Smith (Eds.), *Authorizing readers: Resistance and respect in the teaching of literature* (pp. 88-102). New York: Teachers College Press.
- RAND Reading Study Group. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Arlington, VA: RAND.
- Retelsdorf, J., & Köller, O. (2014). Reciprocal effects between reading comprehension and spelling. *Learning and Individual Differences*, 30, 77-83. <http://dx.doi.org/10.1016/j.lindif.2013.11.007>
- Reznitskaya, A., & Glina, M. (2013). Comparing student experiences with story discussions in dialogic versus traditional settings. *The Journal of Educational Research*, 106(1), 49-63. <http://dx.doi.org/10.1080/00220671.2012.658458>
- Reznitskaya, A., & Gregory, M. (2013). Student thought and classroom language: Examining the mechanisms of change in dialogic teaching. *Educational Psychologist*, 48(2), 114-133. <http://dx.doi.org/10.1080/00461520.2013.775898>
- Reznitskaya, A., Glina, M., Carolan, B., Michaud, O., Rogers, J., & Sequeira, L. (2012). Examining transfer effects from dialogic discussions to new tasks and contexts. *Contemporary Educational Psychology*, 37(4), 288-306. <http://dx.doi.org/10.1016/j.cedpsych.2012.02.003>
- Roswell, F. G., Chall, J. S., Curtis, M. E., & Kearns, G. (2005). *Diagnostic Assessments of Reading (DAR)* (2nd ed.). Itasca, IL: Riverside Publishing.

- Santa, C. M., & Høien, T. (1999). An assessment of Early Steps: A program for early intervention of reading problems. *Reading Research Quarterly*, 34(1), 54-79. <http://dx.doi.org/10.1598/RRQ.34.1.4>
- Schwartz, R. (2006). Literacy learning of at-risk first-grade students in the Reading Recovery early intervention. *Journal of Educational Psychology*, 97(2), 267-267. <http://dx.doi.org/10.1037/0022-0663.97.2.257>
- Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K., Fletcher, J., Brady, S., et al. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading*, 3(1), 69-94. http://dx.doi.org/10.1207/s1532799xssr0301_4
- Stahl, K. (2004). Proof, practice and promise: Comprehension strategy instruction in the primary grades. *The Reading Teacher*, 57(7), 598-609.
- Stanovich, K. (1986). The Matthew Effect: Some consequences of individual differences in the acquisition of reading. *Reading Research Quarterly*, 21(4), 360-407. <http://dx.doi.org/10.1598/RRQ.21.4.1>
- Takala, M. (2006). The effects of Reciprocal Teaching on reading comprehension in mainstream and special (SL) education. *Scandinavian Journal of Educational Research*, 50(5), 559-576.
- Tate, R. L., & Pituch, K. A. (2007). Multivariate hierarchical linear modeling in randomized field experiments. *Journal of Experimental Education*, 75(4), 317-337.
- Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary grade reading instruction in low-income schools. *Elementary School Journal*, 101(2), 121-165.
- Torgeson, J. K., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., Mansfield, W., et al. (2006). *National assessment of Title I interim report to congress: Volume II: Closing the reading gap, first year findings from a randomized trial of four reading interventions for striving readers*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.
- Vadasz, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, 98(3), 508-523. <http://dx.doi.org/10.1037/0022-0663.98.3.508>
- Van Allsburg, C. (1993). *The sweetest fig*. Boston, MA: Houghton Mifflin.
- Van den Branden, K. (2000). Does negotiation of meaning promote reading comprehension? A study of multilingual primary school classes. *Reading Research Quarterly*, 35(3), 426-443. <http://dx.doi.org/10.1598/RRQ.35.3.6>
- Wegerif, R. (2008). Dialogic or dialectic? The significance of ontological assumptions in research on educational dialogue. *British Educational Research Journal*, 34(3), 347-361. <http://dx.doi.org/10.1080/01411920701532228>
- Wells, G. (1993). Reevaluating the IRF sequence: A proposal for the articulation of theories of activity and discourse for the analysis of teaching and learning in the classroom. *Linguistics and Education*, 5, 1-37.
- Wilder, A., & Williams, J. (2001). Students with severe learning disabilities can learn higher order comprehension skills. *Journal of Educational Psychology*, 93(2), 268-278. <http://dx.doi.org/10.1037/0022-0663.93.2.268>
- Wilkinson, I., & Son, E. (2011). A dialogic turn in research on learning and teaching to comprehend. In P. P. M. Kamil, E. Moje, & P. Afflerbach (Eds.), *Handbook of reading research*. Vol. IV (pp. 359-387). New York: Routledge.
- Wilkinson, I. A., Murphy, P., & Binici, S. (2015). Dialogue-intensive pedagogies for promoting reading comprehension: What we know, what we need to know. In L. Resnick, C. Asterhan, & S. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 35-48). Washington, DC: AE-RA.

APPENDIX 1: SAMPLE LESSON PLAN AND REFLECTION

Lesson #: 3

Text: *Tree of Birds*, by Susan Meddaugh (1990). Houghton Mifflin.

Pages to be read: All

Rationale: How rich is this book? What characteristics of this text lend it to thoughtful conversation? To what extent will it lend itself to multiple interpretations?

It's not at all clear why the birds congregate and follow Harry. The ending is a surprise and open to a number of different interpretations.

Leader response: What questions do I have for myself about this text?

Why didn't the birds try to get in before Harry let them in?

Book Orientation: One or two salient points or ideas that need to be introduced in order to get students oriented to the text.

- "Green-tufted tropical"

- Birds flying south

Reflection: How did it go? Narrative Commentary. [Written after the discussion.]

There was little discussion/debate about the meaning of the text in terms of the broader plotline. However, there was discussion on some other points. For example, there was discussion as to how Sally [the bird in the story] had gotten injured. Eduardo suggested that perhaps she had hit a branch, but Maria pointed out that there was reference to a car in the text. This led them all to agree it was a car. Later there was discussion as to why Harry was worried about the cold weather. Juliana and Daniel felt it was because he was worried about the birds outside, while Eduardo and Maria thought he was worried about something happening to Sally. We did not revisit the question explicitly at the end. Finally, there was some disagreement when predicting the ending. Juliana had peeked to the final page and "predicted" the outcome pictured there.

The students said they really liked this text, but next time, I would like to pick a text with a bit more "grist for the mill", as this text felt easy for the group, with not quite enough to chew on.

APPENDIX 2: A SHEP TRANSCRIPT.

This transcript is excerpted from a longer conversation among a group of students had about an Aesop's fable, "The Miller, His Son, and Their Donkey" (Pinkney, 2000). In the fable, a miller and his son continually follow the advice of others as they take their donkey to market to be sold. This culminates in them suspending the donkey from a stick and carrying it. As they are going over a bridge:

The donkey began to kick and thrash. The pole creaked and bent under his weight until it broke with a snap, and the poor donkey tumbled off the bridge into the water. The miller and his son were forced to turn again for home, without the money they had hoped to make, and without the donkey they had before. (Pinkney, 2000, p. 39)

The students in this SHEP group periodically had discussed what was happening throughout their oral reading of the story, and they had become quite interested in the question of *why* the donkey was being sold – perhaps it was mean, or perhaps the miller and his son needed the money, or perhaps they no longer had use for "the skinny donkey." Because the students wanted to explore this further, the discussion facilitator had allowed them to do so, even though it was at best peripherally related to the central moral of the story.

After this final segment of the story was read by one of the students (the moral of the story from the original had been omitted in the version the students saw), the group continued exploring a previously raised hypothesis about the reason for selling the donkey, with Thomas being the first to speak:

Thomas: See, they were poor. That's what I just said. (Refers back to text.) *Without the money they had hoped to make.*

Discussion facilitator (DF): Okay, so....

Thomas: So it sounds like they're poor and they needed the money so they could buy something or something.

Adam: Yeah!

Berta: That's why they're selling the donkey.

Adam: To get more money. But the donkey fell off and it might be dead because// and that's very sad.

DF: Okay.

At this point, another student put forward a new area of discussion, which subsequently became a point of disagreement:

Jenny: I wonder what's going on here because it says... all it says is that (refers to text) *the stick broke and the donkey tumbled into the water.* I want to know what happened to the donkey.

Adam: Yeah.

Thomas: He drowned.

Jenny: You don't know that. You don't know that.

Thomas: He's so big, he's not gonna float. He's gonna drown like a log.

Jenny: You don't know, you don't know how deep the water is. {It could be knee deep. It could be ankle deep.

DF: Is there, is there... well// Hold on, hold on, hold on.

Thomas: It sounds here, it sounds like here, it seems like it's pretty deep or else the donkey wouldn't have drowned or died.

Jenny: It didn't say that the donkey dies or drowned.

Thomas: I know. I'm just saying. It could be very deep and it drowned. I'm not saying that's what happened. I'm just saying...

DF: Hold that thought for just a second. Hold that thought for just a second.

Jenny: It could be pretty shallow.

DF: Jenny. Hold on for just a second. That's...That's a, that's a, that's a, that's a, I would think, a very important question. Is there anything in here in the story that could help us answer that? And maybe there isn't, maybe there isn't, maybe there isn't.

The students were sufficiently animated in their discussion that it took several bids for the discussion facilitator to regain the floor. When he did, he did not try to end the debate, or steer it toward his own conclusion. Instead, because the students were pushing into ever more hypothetical territory, he sought to get the students to continue to ground their claims in textual evidence. The conversation then proceeded as follows:

Adam: I don't think there is.

DF: Okay. So we're left with, maybe he drowned, maybe he didn't?

Adam: These guys were walking to the market to sell the donkey so that they could get a lot of money but instead they, the donkey fell...fell off the bridge and now it's dead and now// It said that the miller// (referred to text) *without the money they had hoped to make and without the donkey they had before* and that means that they were supposed to get money but the donkey fell off the bridge.

Jenny: Well...

DF: Alfredo? (to Jenny) Hold on. Alfredo?

Alfredo: Um, but if the donkey didn't die, why didn't they just get it out?

DF: What do you mean? Explain.