Effects of Teacher-Delivered Book Reading and Play on Vocabulary Learning and Self-Regulation among Low-Income Preschool Children

David K. Dickinson, Molly F. Collins, Kimberly Nesbitt, Tamara Spiewak Toub, Brenna Hassinger-Das, Elizabeth Burke Hadley, Kathy Hirsh-Pasek & Roberta Michnick Golinkoff

Routledge
Taylor & Francis Group

Check for updates

Effects of Teacher-Delivered Book Reading and Play on Vocabulary Learning and Self-Regulation among Low-Income Preschool Children

David K. Dickinson[a], Molly F. Collins[a], Kimberly Nesbitt[b], Tamara Spiewak Toub[c], Brenna Hassinger-Das[d], Elizabeth Burke Hadley[e], Kathy Hirsh-Pasek [c], and Roberta Michnick Golinkoff[f]

[a]Vanderbilt University; [b]University of New Hampshire; [c]Temple University; [d]Pace University; [e]University of South Florida; [f]University of Delaware

ABSTRACT

There is a need for empirically based educational practices shown to support learning, yet validation tends to require a high degree of experimental control that can limit ecological validity and translation to classrooms. We describe our iterative intervention design to support preschoolers' vocabulary through book reading coupled with playful learning, including the process of translating research-based methods to an authentic teacher-delivered intervention. Effectiveness of the teacher-implemented intervention was examined by comparing book reading alone versus book reading plus play in supporting vocabulary development in preschoolers ($N = 227$) from low-income families with diverse backgrounds. Teachers used definitions, gestures, and pictures to teach vocabulary. During play, teachers led play with story-related figurines while using target vocabulary. Ten teachers read books and engaged children in play (read + play [R + P]), and 6 used only book reading (read-only [RO]). For children in both the R + P and RO conditions, within-subjects analyses of gains on taught versus control words revealed large effects on receptive (R + P, $d = 1.08$; RO, $d = 0.92$) and expressive vocabulary (R + P, $d = 1.41$; RO, $d = 1.23$). Read-only had a statistically significant effect ($d = 0.20$) on a standardized measure of receptive vocabulary, but there were no statistically significant differences between conditions. Moderate to large effects were found using an expressive task when words were tested 4 months after they were taught. Implications for curriculum design and the potential benefits of enhancing children's vocabulary through book reading and playful learning are discussed.

Low-income preschool children

Vocabulary in the preschool years is strongly related to later vocabulary and is predictive of reading comprehension throughout the schooling years (e.g., Cunningham & Stanovich, 1997; Dickinson & Porche, 2011; National Institute for Child Health and Human Development Early Child Care Research Network, 2005; Smith, Borkowski, & Whitman, 2008; Storch & Whitehurst, 2002). Given the importance of language to later reading and evidence that children from lower

CONTACT Molly F. Collins molly.collins@vanderbilt.edu Department of Teaching and Learning, Vanderbilt University, Nashville, TN 37203, USA.

socioeconomic backgrounds lag behind in language abilities and early reading success (Chatterji, 2006; National Center for Educational Sciences, 2013), there is a need to identify effective methods for bolstering language. Unfortunately, efforts to use comprehensive full-day interventions often have met with limited success (Dickinson, 2011; Dickinson, Freiberg, & Barnes, 2011; Justice, Mashburn, Pence, & Wiggins, 2008; Powell, Diamond, Burchinal, & Koehler, 2010). Those that have been successful are resource-intensive and involve coaching, professional development, a strong curriculum, and sustained efforts (Wasik, Bond, & Hindman, 2006; Weiland & Yoshikawa, 2013; Wilson, Dickinson, & Rowe, 2013).

Interventions with a singular focus on vocabulary are less demanding of program resources than are full-day curricular interventions, and more often, they have been found to have modest effects (Marulis & Neuman, 2010; Neuman, Newman, & Dwyer, 2011). Book reading is the method most commonly employed for such interventions and is consistently associated with language gains (Mol, Bus, & de Jong, 2009; National Early Literacy Panel, 2009). Efforts that link book reading to supplementary activities such as science (Gonzalez et al., 2010; Neuman & Dwyer, 2011) and play (Toub et al. (2018); Weisberg et al., 2015) have resulted in somewhat larger gains in vocabulary than those that employ only book reading. This finding suggests the potential for increased word learning when book reading is augmented by activities that support vocabulary.

Despite mounting evidence of the efficacy of book reading combined with additional activities, no widely employed scalable methods are available for early childhood teachers to use. This article includes the results of an effort to create a scalable intervention that links book reading to teacher-supported play for use in early childhood classrooms.

Language intervention strategies and challenges

Decades of research on language learning in homes and classrooms have identified evidence-based strategies for supporting language, many of which are used during teacher-led book reading and teacher-supported play. For example, effective principles of language development include repeated exposure, active involvement of children that includes opportunities to use words, and explicitly telling children the meanings of words (Biemiller & Boote, 2006; Harris, Golinkoff, & Hirsh-Pasek, 2011; Hirsh-Pasek & Golinkoff, 2012). In addition, as children hear words used in meaningful and varied contexts, they build conceptual information and associations and acquire syntactic information about words (Hassinger-Das, Toub, Hirsh-Pasek, & Golinkoff, 2017a). Such contextually embedded exposure supports initial word learning during book reading even in the absence of explicit definitions (Dickinson et al., in press).

Book reading

Book reading is a recognized venue for activating these principles. Research has shown that teachers are most likely to engage in talk about words during this common activity (Dickinson, Hofer, Barnes, & Grifenhagen, 2014). Further, a meta-analysis of predictors of language development conducted by the National Early Literacy Panel revealed that shared book reading is the intervention associated most often with beneficial effects (National Early Literacy Panel, 2009). Other meta-analyses have also shown that book reading is commonly associated with language learning (Mol & Bus, 2011; Mol et al., 2009). Classroom-based interventions that have been found to have beneficial effects on language typically include book reading as a key element (Gonzalez et al., 2010; Neuman & Dwyer, 2011; Neuman et al., 2011; Wilson et al., 2013).

Play

Playful learning (Hirsh-Pasek, Golinkoff, Berk, & Singer, 2009) is another approach to teaching vocabulary that taps into these principles while requiring more activity on the part of the child. Two types of playful learning are free play and "guided play" (Fisher, Hirsh-Pasek, Golinkoff, Singer, & Berk, 2011; Hassinger-Das et al., 2017b; Weisberg, Hirsh-Pasek, Golinkoff, Kittredge, & Klahr, 2016). In free play, children engage in playful activities autonomously without adult involvement or imposed structure. They organically stumble upon learning opportunities, though not necessarily the same ones that adults would prioritize. We have found that such play does not contribute to learning of specific target vocabulary (Dickinson et al., in press). Another type of playful learning, "guided play," refers to learning experiences that combine the child-directed nature of free play with a focus on learning outcomes and gentle adult mentorship. The adult may set up scenarios to encourage children to engage in experiential learning with goal-related materials as in Montessori programs (Lillard, 2016) and encourage them to actively participate during the play. Through guiding questions and commentary, they help children explore in goal-related ways without removing children's control over the play.

Incorporating play into vocabulary may support learning and transfer because playful situations are meaningful to children and provide opportunities for children to rehearse what they have learned. Roskos and Burstein (2011) supplemented instruction during book reading with play and found it resulted in significant gains on the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007). Han, Moore, Vukelich, and Buell (2010) worked with 4- to 5-year-old high-risk Head Start children (i.e., PPVT scores more than one standard deviation below national means) and compared two approaches: vocabulary instruction only during book reading and instruction during book reading and play that included use of story-based replicas and scripted adult support. After 4 months, both conditions resulted in statistically significant growth on the PPVT with no statistically significant differences between groups. However, 62% of the children in the play condition moved into the "age-appropriate" standardized score range of 85 to 115 on the PPVT (Dunn & Dunn, 2007) compared with only 44% in the book reading-only condition. This finding is encouraging, but even though children were randomly assigned to condition, it was not clear how closely matched the control and experimental children were at pretest or how many started out close to the 85th percentile.

Play may facilitate deeper learning than is gained through book reading alone. During play, teachers may be more able to engage in responsive conversations (Hadley, 2017). Also, there may be greater potential for generalization of word knowledge when exposure to words occurs across more than one context. Helping children acquire and retain deep knowledge of words is valuable because depth of knowledge plays a role in predicting reading comprehension over and above breadth of knowledge (Leider, Proctor, Silverman, & Harring, 2013; Proctor, Silverman, Harring, & Montecillo, 2012). Depth of knowledge is fostered by supplying conceptual information about words, having children talk about word meanings, and encouraging children to use words in varying contexts (McKeown & Beck, 2014).

In our approach, we sought to deepen learning by having teachers guide children in book-related play that incorporated new vocabulary. By asking questions and prompting children to use and reflect on the words, teachers can encourage deeper learning. A previous

study using approaches similar to those we employed compared word learning through book reading followed by play versus book reading followed by review using picture cards. Play resulted in a greater depth of knowledge as revealed by stronger performance on a production task in which children defined the taught words (Dickinson et al., in press) using a measure that credited children for multiple types of knowledge (Hadley, Dickinson, Hirsh-Pasek, Golinkoff, & Nesbitt, 2016).

Previous language

In language-focused interventions, children with stronger language often make the greatest gains in vocabulary (Blewitt, Rump, Shealy, & Cook, 2009; Cain & Oakhill, 2011; Penno, Wilkinson, & Moore, 2002; Silverman & Crandell, 2010), a phenomenon referred to as the Matthew Effect (Stanovich, 1986). Sometimes high levels of explicit and sustained instruction may reduce or eliminate these effects (Hindman, Wasik, & Erhart, 2012; Loftus, Coyne, McCoach, Zipoli, & Pullen, 2010; Neuman & Dwyer, 2011; Penno et al., 2002; Pollard-Durodola et al., 2011). Indeed, in some cases, children with lower initial levels of vocabulary make greater gains (Hassinger-Das et al., 2016; Justice, Meier, & Walpole, 2005). However, successfully countering Matthew Effects is a two-edged sword. Matthew Effects may be eliminated by placing a ceiling on the growth of children with more language skill by teaching very few words or by continuing instruction longer than necessary for advanced learners. We sought to meet the needs of both less advanced and more advanced language learners by supplying repeated, explicit, and engaging instruction while teaching 16 words per book—more than have been taught in many book-based interventions. We hoped to provide enough opportunities for more advanced word learners to be challenged while giving sufficient support to enable those with lower initial levels to learn words as well.

Fidelity of implementation

A reason why interventions may fail to have much impact is because it is hard for teachers to make sufficient adjustments in their language use to make a difference in children's language learning (Dickinson, 2011). We hypothesized that we could help teachers significantly enrich their support for language by giving them materials and prompts that would help them draw children into productive interactions during book reading and play. Book reading is a promising setting because at that time, teachers naturally tend to engage in more language-supporting talk than in other classroom settings (Dickinson et al., 2014). During play, teachers were working with a small group and could attend to and foster individual children's contributions. Given these findings, we used book reading supplemented by play and created materials for teachers that helped structure their adoption of recommended methods of reading and playing. We also provided modest levels of coaching. We tracked the fidelity with which they implemented our methods because if we failed to find effects of our approach, weak fidelity to our methods could account for our findings. If we did find effects, a positive association between fidelity and outcomes would further constitute additional evidence of the association between our approach and the observed benefits.

Self-regulation

Self-regulation, or the ability to control one's thoughts and behaviors in service of goal, has also been found to be predictive of concurrent and long-term academic success (Duncan et al., 2007; Fuhs, Nesbitt, Farran, & Dong, 2014; Nesbitt, Farran, & Fuhs, 2015; Smith et al., 2008). Previous research conducted with preschool-aged children has demonstrated associations between improvements in language learning and growth of self-regulatory abilities (Bohlmann, Maier, & Palacios, 2015; Dionne, Tremblay, Boivin, Laplante, & Perusse, 2003; Hammond, Muller, Carpendale, Bibok, & Liebermann-Finestone, 2012; Landry, Miller-Loncar, Smith, & Swank, 2002). Preschool classrooms can contribute to growth in self-regulation. Nesbitt et al. (2015) assessed language and literacy in the fall and evaluated self-regulation using direct assessments and teacher ratings. They observed children's participation in classroom activities and found that self-regulatory ability predicted classroom participation, which mediated effects on fall-to spring language and early reading gains. Attending preschool classrooms with a strong curriculum also has been found to be associated with gains in self-regulation (Weiland & Yoshikawa, 2013).

Self-regulation and language may be related because language can be used as a tool to control one's actions and thoughts (Vygotsky, 1978). Support for this theory has come from studies that have shown associations between use of private speech and tasks that require focused attention (Berk & Spuhl, 1995; Bivens & Berk, 1990; Fernyhough & Fradley, 2005). Learning words that refer to mental states was associated with improvement in self-regulation (Carlson, Mandell, & Williams, 2004), and book discussions often explored character motivations and feelings, thereby providing opportunities to learn these terms (Dickinson et al., in press). A third reason for the association between language development and self-regulation may be that adult–child interactions that foster development of one ability also foster growth of the other. Landry and colleagues (Landry, Smith, Swank, & Miller-Loncar, 2000) followed children from age 2 years to 4.5 years and found that early parental support for maintaining attention was related to self-regulatory ability and language. Bindman and colleagues reported similar results in a study of children ranging in age from 3 years to kindergarten age (Bindman, Hindman, Bowles, & Morrison, 2013).

We hypothesized that classroom activities that provide routine opportunities for sustained engagement could likewise foster improved self-regulation. Book reading is one such activity. Books are read in an engaging manner, and teachers encourage children to sustain attention throughout the activity. Play might also yield such benefits because it can create opportunities for deep, sustained, and self-direct engagement (Weisberg et al., 2015) as well as opportunities for teachers to encourage language use and for children to use new vocabulary.

Current study

Here, we report results from the final year of the Read-Play-Learn (RPL) Project and describe the 2-year process of creating, testing, and refining methods for preschool teachers to use to teach vocabulary through book reading and teacher-guided play. During the 2-year development process, empirical findings and feedback from teachers, our advisory team, and instructional coaches resulted in our methods shifting from tightly controlled tests of methods delivered by research personnel to teacher-delivered interventions.

In the 3rd year, we tested the impact of 12 weeks of teacher-delivered use of the two intervention methods we devised, book reading and teacher-guided play, on children's growth in vocabulary and self-regulation. To test if the combined effects of book reading and play were more effective than book reading alone, we had two conditions: reading only (RO) and reading plus play (R + P). We examined the following research questions:

(1) Do children in the two intervention conditions make significant pretest to posttest gains in knowledge of vocabulary explicitly taught, and are these gains greater than those in knowledge of nontaught control words? Are significant pretest to posttest gains seen for general receptive vocabulary and self-regulation as well?
(2) Is there greater improvement in taught vocabulary knowledge for children in the R + P condition than for children in the RO condition? Are gains in vocabulary knowledge retained after an extended delay, and does retention vary by condition?
(3) Did teachers deliver the interventions with fidelity, and were there particular elements of the intervention more or less frequently implemented?

Method

Participants

Participants were recruited from 10 classrooms that were part of Tennessee's Voluntary Pre-K Program in a major city and 6 Head Start preschool classrooms in Pennsylvania (see online supplement Table 1).All classrooms served low-income populations. Forty-two percent of children were African American, 32% were Hispanic/Latino, 15% were European American, and 9% were Multiracial/Other. Thirty-six percent of parents reported that a language other than English was spoken in the home. All teachers held a bachelor's degree, and nine held master's degrees. The average number of years of teaching experience in the field of early childhood education was 15.4 years (range = 1–40 years). The same classrooms participated throughout the year, and fall and spring samples were very similar.

Teachers in Tennessee's participating pre-K classrooms were instructed to distribute consent forms to the guardians of children who had not been referred for individualized education programs in areas such as intellectual or language delay, who had no significant visual or hearing impairments, and who demonstrated sufficient English proficiency for participation. At the Pennsylvania Head Start site, teachers distributed consent forms to

Table 1. Fall phase receptive and expressive vocabulary knowledge means (standard deviations) by condition and level of instruction.

	Read + Play (n = 129)			Read Only (n = 81)		
	Pretest	Posttest	Delay	Pretest	Posttest	Delay
Receptive Vocabulary						
Taught: Percent Correct	0.40 (0.11)	0.57 (0.15)	0.35 (0.15)	0.41 (0.12)	0.54 (0.15)	0.37 (0.14)
Control: Percent Correct	0.37 (0.19)	0.39 (0.20)	0.26 (0.17)	0.38 (0.20)	0.39 (0.18)	0.31 (0.18)
Expressive Vocabulary						
Taught: Average Points per Word	0.21 (0.17)	0.48 (0.35)	0.26 (0.18)	0.22 (0.20)	0.44 (0.28)	0.25 (0.19)
Control: Average Points per Word	0.11 (0.14)	0.11 (0.17)	0.09 (0.13)	0.11 (0.17)	0.13 (0.13)	0.12 (0.16)

Note. Pooled means and standard deviations across 10 imputed data sets.

guardians of all children, except those who had participated in a previous phase of this project. After consents were collected, children in those classrooms were approved for full participation if they met the aforementioned criteria.

In the fall, consent forms were obtained for 234 children across both sites. Seven children withdrew from the study before data collection, and 17 were excluded from analyses based on developmental delays. Thus, the final sample of children who participated in the fall phase was 129 children in the R + P condition and 81 children in the RO condition (see online supplemental Figure 1 for the study consort chart). Three children in the RO condition and 2 children in the R + P condition left the preschool programs between the fall and spring semesters. At the onset of the spring semester, an additional 15 consent forms were received for the R + P condition, and an additional 6 were received for the RO condition. Two children who participated in the spring were excluded based on developmental delays, and 2 were excluded due to administration error; thus, in the spring, 138 children participated in the R + P condition and 84 participated in the RO condition. All children with complete data on at least one measure were included in the analyses, and on average, 9 children per classroom were in the analytic sample. Across phases and conditions, approximately 52% of children were female, and children's mean age was 4;4 in the fall and 4;9 in the spring. See online supplemental Table 1 for full descriptives by condition and phase.

Average attendance for the fall theme (Theme 1) was 3.50 of the 4 reading sessions provided per book (SD = 0.80, range = 0–4 days) and 2.61 of the 3 play sessions (SD = 0.73, range = 0–3 days). Average attendance for the spring theme (Theme 2) was 3.69 of the 4 reading sessions provided per book (SD = 0.62, range = 0–4 days) and 2.82 of the 3 play sessions (SD = 0.44, range = 0–3 days).

Developing the intervention

The RPL Project spanned 3 years. The first 2 years were dedicated to developing and refining strategies. In Year 3, we tested our methods. In the first 2 years, we created our methods, tested to see if the methods were effective when used with high fidelity, and made adjustments as we turned the instruction over to teachers. Our goal was to ensure usability in preschool classrooms while retaining the effectiveness that we obtained when using our controlled form of delivery by trained intervention specialists (ISs). We were faced with a multitude of practical challenges, and as we grappled with them, we acquired considerable practical knowledge related to delivering interventions in classrooms. Our efforts also led us to try out different approaches to teaching that shed new light on vocabulary acquisition in classrooms. One aspect of vocabulary acquisition that we learned about was the amount and type of instruction for different types of words that is needed to foster learning. A related issue was the number of words that could be taught and learned during an instructional event. These issues were forced on us as we dealt with time constraints. There was limited time to teach words during group book reading due to competing demands on teachers' time and children's attention spans. Time was a particularly pressing matter for our play methods because it required teachers to work with small groups of children across several sessions. Finding time for three or four 12-min to 15-min play sessions with groups of three to five children proved to be a challenge for many.

Instructional intensity

In our efforts to deal with issues related to instructional intensity during book reading, we varied the number of words taught during a reading by teaching 10 words in one round and 16 in others. We also varied the number of times we supplied explicit instruction about word meanings. In one round of our testing, we provided instruction four times, and in another iteration, it was done twice. We were heartened to find that children learned roughly as many words when we introduced 16 new words as when we taught 10 and that direct instruction during two sessions was sufficient if the words also received passing attention on two other occasions. Thus, in our final design for each book, we taught 16 words divided into two sets. Set A received direct instruction during the first and third reads, and Set B received direct instruction during the second and fourth reads. Incidental instruction was given during the readings when no explicit teaching occurred.

Time constraints

The time constraints were even more of a challenge for play. Not only was it hard to find classroom time for the teacher to work with multiple small groups, but also, they needed repeated opportunities for play related to a story so they could begin using the words being taught. Time constraints were exacerbated by the tension between our instructional objective of teaching a specific set of words and our desire for children to engage in child-initiated play. Children who are just learning new words, especially those for whom English is a new language, are often reluctant to begin using unfamiliar words. It takes a skilled teacher to entice children to use such words. Furthermore, play can move in unexpected directions and make it hard for teachers to ensure that they use the specified words often enough. Thus, ensuring sufficient word exposure and instruction during play was particularly challenging. The number of play sessions in which children participated varied from three to four. In one iteration, we taught 10 words across four sessions, with half the words receiving direct instruction each day. In a second iteration, we taught 16 words divided into two sets. In one iteration using that design, our ISs led the groups, and in a second round, teachers with strong coaching support delivered the instruction. Children displayed evidence of large growth ($d > 1.0$) in receptive and expressive knowledge for both studies (Dickinson et al., in press). Thus, across studies and intervention methods, we found that children could learn up to 16 new words while receiving explicit instruction on two occasions and incidental reinforcement the other two times.

Nature of instruction

A second set of issues born of pragmatic concerns related to the extent to which we used explicit instructional methods. Our need to teach words efficiently pushed us toward using explicit methods found to be effective previously (Biemiller & Boote, 2006; Coyne, McCoach, & Kapp, 2007; Coyne, McCoach, Loftus, Zipoli, & Kapp, 2009). However, explicit teaching can be time-consuming, may bore children, and tends to break the flow of an activity—book reading or play. As we devised our book-reading method, we tested two explicit methods: formal word definitions alone and definitions supplemented with conceptual information. We reasoned that added knowledge would help children gain deeper knowledge and might reduce boredom. A third novel approach was tried because it mimicked a common classroom strategy. Teachers used the target words as they briefly recalled and explained what happened in a story event. Our thinking was that enhanced event comprehension combined with hearing the word used in context might

boost word learning. Somewhat to our surprise, we found that all three methods resulted in significant and modest-sized growth in word knowledge and no statistically significant differences by method (Dickinson et al., in press). The finding that using words when recalling and explaining events resulted in growth in word knowledge suggests that word meaning may be fostered by associating words with salient events. In practical terms, it opens another avenue for vocabulary instruction and suggests that intentional use of novel words while discussing material may augment word learning. We used elements of all three of our vocabulary instruction methods when creating the approach to instruction used in the pilot study.

Another pragmatic concern that resulted in use of a method with a conceptual payoff was our need to ensure reasonably similar and concentrated instruction about word meanings across methods and teachers. Teachers vary in their ability to use more complex methods of play and book reading, and all teachers lead busy time-pressured lives. Also, our methods varied in the extent to which we could be certain that each word was taught effectively. To address these concerns, we introduced the use of picture cards along with iconic gestures prior to each activity. Teachers achieved a high degree of fidelity in their use and reportedly liked having them as supports. We did not experimentally control the use of pictures with gestures, but the potency of our methods and the similarity of effects across play and book-reading approaches suggest that providing iconic and gestural supports for word meaning may play an important role in enriching children's semantic representations.

Retention of learning

A question of great importance to both educators and psychologists is the extent to which children remember what has been taught. Studies of word learning have rarely examined long-term retention despite its obvious importance for understanding learning mechanisms. Examining word retention is critical, because sustained recall helps determine whether children are likely to derive long-term benefits from instruction. Research has suggested that children with lower initial language and literacy skills are less likely to retain word knowledge over time compared with their higher-achieving peers (Pullen, Tuckwiller, Konold, Maynard, & Coyne, 2010). We were particularly interested in determining the retention rates of the children in our study population due to their at-risk status. As such, to investigate knowledge retention, we conducted a delayed evaluation of word learning.

Our final instructional design was the byproduct of substantial and continuous refinement based on feedback from seasoned educators. We collaborated with an advisory team of highly skilled preschool teachers during the 2nd year, garnered feedback from participating teachers in multiple focus groups, and had ongoing discussions with our instructional coaches who were experienced educators. Insights from these professionals played a pivotal role in shaping our design of instructional materials.

Pilot study design

Methods

Classroom teachers delivered the intervention during September to March. There were two themes (dragon, farm) and two books in each theme. One theme was delivered October to November, and the other was delivered in February to March. The orders of themes and books within themes were counterbalanced. Ten classrooms were randomly

assigned to the R + P condition in which the teacher read thematically related books and led play groups of three to five children with toys related to the current book. Six classrooms were randomly assigned to the RO condition in which the teacher only read books. At one site, classrooms were assigned at the school level (three clusters of 2 classrooms: 4 R + P, 2 RO), and at the other site, 10 classrooms were assigned at the classroom level (6 R + P, 4 RO). The research team and school administrators asked all teachers to not discuss their activities with other teachers. Moreover, the potential for influence by the RP teachers on the RO teachers was limited further as the RO teachers did not receive play training, guidance materials, or toys.

Book and word selection

For each of the two themes (dragon and farm), we used two core books and four supplementary theme-related books. All core books were comparable in terms of the pictorial representations of taught words, text complexity, and length. There were 16 taught words per core book, representing a mix of verbs and abstract and concrete nouns. Words were selected using the following procedures to ensure that the words were sufficiently difficult to be unknown to participants at baseline. First, we identified Tier 2 words (Beck, McKeown, & Kucan, 2002). If a book did not contain 16 Tier 2 words, we supplemented the text while taking care not to change the storyline. We cross-referenced our words with Biemiller's (2010) list of words; 10 taught words did not appear on the list. Of the 54 taught words that were on the list, 91% were characterized as at least Level T2— high-priority words that are typically known by more advanced students by the end of second grade and are not known by at-risk students. According to Snow's list of rare words (Dickinson & Tabors, 2001) that was based on Dale-Chall's (Chall & Dale, 1995) list of common words, 66% of our taught words were rare. We ensured that selected words could be explained in child-friendly terms and were semantically and phonologically distinct. Control words were selected from those used in the previous 2 years of the study using the same procedures as our taught words. They were matched on word type (e.g., form class and imageability) to taught words.

Book reading

Children in both conditions heard four books during the course of 12 weeks (3 weeks per book). Each book was read four times. To keep the teaching and learning load manageable, the 16 taught words were split into two groups, with one half serving as the "focus words" on Reading Days 1 and 2 and the other half serving as focus words on Reading Days 3 and 4. "Focus words" received rich explanation (described in the next paragraph) while the remaining 8 words were read aloud as they appeared in the book but were not discussed during the session.

Focus words were introduced before the reading using picture cards that depicted the words with images different from those in the book. Teachers used our child-friendly definition of the word, showed the picture card, said the word (or had children recall it, in later sessions), and had children repeat the word. Teachers were encouraged to use gestures and have children gesture (e.g., mime *rummaging* with their hands). During the reading when focus, taught words were encountered, teachers pointed to the relevant picture, defined the word, asked children to say the word, and encouraged them to gesture. When words were new, teachers prompted children to repeat the word and

gesture after them. In later readings, teachers paused and encouraged the children to "fill in the blank" for a word as they read the book or after they gave the definition and gesture. For all readings, teachers used focus words as they discussed events in the story: "The dragon has some family members who lived a long, long time ago. Those family members are his *ancestors*, and they left those books for him to read."

To support story understanding, teachers asked scripted discussion questions during the readings. For the first two readings, prompts probed literal details and some inferential content. During the third and fourth readings, prompts mostly requested inferential responses.

For each core book, teachers were given two supplementary books matching the theme and were asked to read each one twice during the course of the 3 weeks. We provided some suggested vocabulary words, but there was no standard protocol. Also, teachers were allowed to review the picture vocabulary cards from the core books' intervention materials at other times during the associated weeks. Additional details regarding the conceptualization of the book-reading intervention are provided in Dickinson et al., (in press).

Play

Children in the R + P condition participated in three adult-directed play sessions related to each book. As the goal of the play session was to provide opportunities for children to use words in a meaningful context in accordance with their play ideas, we incorporated techniques that encouraged children's use of the words.

As with the book reading, a subset of taught words was designated as focus words for each day of play. Each taught word was reviewed in a preplay story review on either Day 1 or Day 2 and then during the play itself on 1 of the 3 days. On Days 1 and 2, prior to starting the play itself, teachers were given illustrations from the book to use in a brief review of the story and taught words. Eight taught words were reviewed on Day 1, and the other eight words were reviewed on Day 2. Teachers were given guidance cards that outlined ways to discuss what was happening in each picture, use each word, provide the definition, and ask children to say it. The use of gestures was encouraged. To provide opportunities for children to use words in their play, we incorporated techniques that have been found to be useful in previous word-learning research (McKeown & Beck, 2014) as well as new research on play (Dickinson et al., in press).

After this review, the teacher led children through a reenactment of the story using our set of figurines (e.g., king, dragon), background pieces (e.g., castle), and replica props (e.g., handkerchief, egg). Teachers were instructed to describe and repeat children's play actions, elaborate children's talk, prompt their word use, and be playful. Guidance materials outlined three major scenes and example language for teachers to use to scaffold the play and incorporate strategic review of the day's designated taught words (i.e., five words per day for Days 1 and 2 and six words on Day 3). For the focus words, teachers were asked to use the word, invite children to say it, ask closed-ended questions (e.g., "Is *mayhem* calm or a little crazy?"), and ask open-ended questions (e.g., "How is *mayhem* different than calm and peaceful?"). Examples of each of these review techniques were provided on the guidance cards. Optional activities were defining the word, gesturing, and inviting children to gesture.

On Day 3, before play began, rather than reviewing the story, the teacher briefly introduced words while handing out the toys (e.g., "You get to be the dragon. Here [pointing] are his *nostrils* and sharp *talons* ….."). Teachers asked children to select a new

topic for their play (e.g., beach, birthday party, picnic at a playground). We supplied framing questions suggesting events that the teacher could use to help get children started. Teachers could also reference their guidance cards for ideas on how to incorporate the six focus words into the play. These guidance cards included examples of closed- and open-ended questions to use. Additional details regarding the conceptualization of the play intervention are provided in Dickinson et al. (in press).

Coaching

Research team members called ISs acted as coaches to assist teachers in adopting the reading and play approaches. The amount of coaching provided was decreased during the year. For Book 1, coaches observed three or four readings and each type of play (reenactment on Days 1 and 2 and new scenario on Day 3). For the remaining books, coaching decreased from two readings and a minimum of two sessions of each type of play for Books 2 and 3 to one reading and one play session.

During visits, the IS used coaching forms that tracked the use of prescribed strategies and taught words. They gave immediate feedback during a brief discussion after the visit (Site 1) or in an e-mail to the teacher (Site 2). Feedback was concise and generally focused on one or two strengths and one or two areas for improvement. End-of-week feedback was given through a conversation (Site 1) or through an e-mail (Site 2), with support tapering to "as needed" as the year progressed. Feedback discussed growth in using strategies, children's progress, and planning for the next week. Intervention specialists also completed end-of-week surveys for the research team, which summarized teachers' adoption of all materials, progress in implementing strategies for book reading and play, overall uptake of guidance, and need for additional coaching.

Child measures

Trained assessors individually evaluated children at pretest and posttest. Assessments of taught and control words for Theme 1 were in September and November/December, and for Theme 2, they were held in January and March. The delayed posttest for Theme 1 was in April.

Receptive test

Children's receptive knowledge of taught and control words was tested with a picture vocabulary test similar to the PPVT (Dunn & Dunn, 2007). Each array included the correct referent, a conceptual foil, and a thematic foil (e.g., for the word *throne*: throne, folding chair, and crown, respectively). The images differed from the books' illustrations and the picture vocabulary cards used during the intervention. The experimenter asked, for example, "Where's *throne*?" or "Can you show me *throne*?" and the child's image selection was noted. The task began with four training items, followed by the taught and control words. One easy filler item was to provide successful experiences. Children's scores reflected the percentage of test items (excluding training and filler words) for which they chose the correct referent image. Cronbach's alpha estimates of internal consistency were .77 and .74 for the dragon and farm themes, respectively.

Expressive test

We used the New Word Definition Test-Modified (Hadley et al., 2016). Children were asked to explain the meaning of vocabulary words using words and/or gestures through a procedure modeled after the tool used by Blewitt et al. (2009). Initial prompts included, for example, "What are *talons*?" or "What does *intelligent* mean?" After the child's initial response, we prompted for additional information (i.e., "Can you tell me or show me anything more about *intelligent*?"). Responses were video- or audio-recorded. Cronbach's alphas were .92 and .91 for the dragon and farm themes, respectively.

For both the receptive and expressive tests, children were tested on a subset of taught and control vocabulary. Words were divided into three sets per theme (see *Planned missing data design for vocabulary measures* for a description of this process); each child was tested on two sets per test, and children were randomly assigned to receive a given set of words. Each receptive test set consisted of 14 words to 15 words including 10 to 12 target words, 2 to 3 control words, and 1 easy filler word. Expressive test sets each contained 14 words including 10 to 11 target words, 2 to 3 control words, and 1 filler word.

Coding

Our scoring for the Expressive task was based on Blewitt et al. (2009) and is described in more detail in Hadley et al. (2016). Children's responses were coded for the number of correct *information units*. They received credit for whether a response revealed understanding through synonyms, antonyms, part–whole relations, categorical hierarchical relations, function, perceptual features, context, or gestures. For example, for the word *handkerchief*, a child could say, "Wipe your nose, use for your tears, made out of cloth." This response would receive 3 points for including two functions and one perceptual feature. Scores were the average of the number of information units for items. Before beginning, coders achieved 90% reliability with a gold-standard coder. A gold-standard coder double-coded a random 20% of the assessments with those gold-standard tests being given to coders throughout the coding period to protect against rater drift. The average percent agreement among all coders for all participants was 95.3% (kappa = .84).

Planned missing data design for vocabulary measures

We tested 79 words (40 for the dragon theme, 39 for the farm theme). To reduce potential fatigue, practice effects, and measurement error that could occur from requiring children to complete all possible items, a three-form planned missing data design was employed (see Graham, Taylor, Olchowski, & Cumsille, 2006) to assess children's learning of 32 taught words and 7 to 8 control words per theme. The three-subset design allowed us to cut the number of items given to a child by 33%. Items were divided into three sets, and children were randomly assigned two sets. Each set included 14 items (one dragon item set included 15 items) that were selected so that each set included an approximately equal balance of 1) taught, control, and filler items; 2) items from the two books within the theme; and 3) items across the three form classes (concrete nouns, abstract nouns, and verbs). The sets assigned to a child were the same for the receptive and expressive knowledge measures. The order in which sets were given was counterbalanced across measures. Moreover, the sets varied randomly across the testing periods (i.e., pretest, posttest, delayed test). Because sets were randomly assigned to children, data were

assumed to be missing completely at random, and modern missing data approaches were implemented to account for missingness. Ten multiple imputation data sets were estimated in Statistical Package for the Social Sciences (SPSS) Version 22 (IBM, 2013). All analyses of receptive and expressive data were conducted for every imputed data set, and results were pooled using Rubin's rules (Rubin, 1987), which accounted for variance within and between imputations. Full data were collected from all children for self-regulation assessments and PPVT.

General vocabulary

The PPVT-Fourth Edition (PPVT-4; Dunn & Dunn, 2007) requires a child to point to one of four pictures that matches the word spoken by the examiner. Test–retest reliability coefficients ranged from .92 to .96. The PPVT-4 was administered at pretest of Theme 1 (September; or at pretest of Theme 2 in January for children new to the study) and in early April following the posttest of Theme 2 (March) and the administration of the delayed test of Theme 1 (early April).

Self-regulation

Aspects of children's cognitive self-regulation, namely their ability to inhibit a prepotent response, was assessed using the peg-tapping task (Diamond & Taylor, 1996). On this task, children tap their peg twice after the interviewer taps once, and vice versa. After initial practice items and rule checks, there were 14 test trials. This task was administered at the same time as the PPVT.

Additionally, teachers rated aspects of children's behavioral regulation observed in the context of the school day using the Cooper-Farran Work-Related Skills (Cooper & Farran, 1988). This test consists of 16 items about independent work, following instructions, and completion of games and activities. Teachers completed these forms for each child at the start of the child's participation (i.e., September for Theme 1 or February for children starting during Theme 2) and at the end of Theme 2 (i.e., March).

Teacher measures

We collected information about teachers prior to and during implementation.

Teacher professional experience questionnaire

Prior to the intervention, teachers completed a questionnaire about education, years of experience, coursework in professional preparation programs, and in-service professional development opportunities.

Book-reading fidelity

For each of the 16 classrooms, we coded fidelity of implementation from videos of two readings from Book 2 (i.e., Reading 1 or 2 and Reading 3 or 4) and Book 4 (i.e., Reading 1 or 2 and Reading 3 or 4). The videos were coded for use of nine behaviors. Scores were averaged across the two readings per semester to yield a fall score and a spring score. Overall Reading Fidelity was an average of the fall and spring scores. Coders were trained, and videos were randomly selected within Readings 1 and 2 and Readings 3 and 4.

Approximately 25% of the videos were double-coded. Overall agreement between coder and master coder was 83% or greater for all coders.

Play fidelity

A representative subset of four videos from each R + P classroom's play sessions was coded for play fidelity. The sample typically consisted of two videos from Book 2 and two videos from Book 4, with the two videos per book split across days to capture both types of play sessions. Videos included both the preplay vocabulary and story review (Days 1–2) and the play session itself (Days 1–3).

We sought to document both the degree to which teachers adhered to our vocabulary review procedures before and during play and the degree to which they used the desired play style. The coding scheme captured the proportions of the day's assigned words for which the teacher displayed each review behavior (e.g., used the word, defined the word, invited children to say the word). We coded their fidelity to our methods in the preplay word review (Days 1– 2) and during the play itself (Days 1–3). Average proportions were calculated for each classroom across the four videos.

To capture teachers' adherence to the desired play style, we coded the proportion of 1-min segments of the play session during which the teacher showed specific behaviors characteristic of our play approach (e.g., joined children's play by becoming a character, built on children's contributions), as opposed to a free-play or overly directive approach.

Each of two trained play fidelity coders served as the main coder for half the videos, and 20% of the sample was blindly double-coded to check reliability. The eight-video subsample was randomly selected within constraints that ensured a spread across combinations of books and play days. Agreement on the double-coded subsample ranged from Kappa = .64 to Kappa = .92 on the items used in our analyses.

Results

Evaluation of intervention conditions

Analytic approach

To examine the impact of the studies' two intervention conditions, a series of multilevel analyses was completed to account for repeated assessments within children and nesting of children in classrooms; we also accounted for the assignment of experimental condition to the classroom. First, to test if children made significant pretest to posttest language and self-regulation gains in each intervention condition, we conducted within-subject (i.e., repeated assessment) comparisons of pretest, posttest, and delayed posttest scores. Further, to explore if gains for taught words were larger than control words, within-subject comparisons were conducted to contrast individual children's gains on both taught and control words. Lastly, to examine if RO classrooms made differential vocabulary and self-regulation gains compared to R + P classrooms, between-subject comparisons were made that adjusted standard error to account for the intervention being delivered at the classroom level.

Evidence of learning

To determine whether children learned taught words better than control words, we first tested to see if children made significant gains from pretest to posttest. Across both

conditions, children made large and significant pretest to posttest gains on taught voca-bulary words but not control words (see Table 4). For taught words, growth in receptive knowledge was large for both R + P ($d = 1.32$) and RO ($d = 1.02$). Growth was similarly large for expressive knowledge (R + P, $d = 1.12$; RO, $d = 0.94$). For control words, growth was minimal for receptive knowledge (R + P, $d = 0.11$; RO, $d = 0.04$) and absent for expressive knowledge (R + P, $d = 0.10$; RO, $d = 0.01$).

Within-subject contrasts between taught and control words revealed large effects of both conditions in the fall and spring. Multilevel models were conducted separately by condition with level of instruction repeated across children who were nested within classrooms. Site, pretest, age, gender, and second language (language other than English spoken at home) were included as covariates. Collapsed across phase (i.e., fall, spring) and theme (i.e., dragon, farm), we found large effects ($p < .001$) for both conditions for receptive knowledge (R + P, $B = .191$, $SE = .017$, $d = 1.08$; RO, $B = .153$, $SE = .019$, $d = 0.92$) and expressive knowledge (R + P. $B = .390$, $SE = .026$, $d = 1.41$; RO, $B = .267$, $SE = .026$, $d = 1.23$).

We tested for evidence of growth in generalized vocabulary knowledge using age-normed scores from the PPVT (see Table 3 for descriptives). We found that children in the RO condition made small but educationally significant gains ($d = 0.20$). This finding was not observed in the R + P condition ($d = 0.04$). We also tested for growth in self-regulation. Across both conditions, children made significant gains on our direct assess-ment through peg tapping (R + P, $d = 0.24$; RO, $d = 0.37$) and our teacher-reported measure (R + P, $d = 0.29$; RO, $d = 0.44$). As these tools were not age-normed, we could not disentangle age-related effects from possible intervention effects.

Condition effects

Classrooms were randomly assigned to treatment, but given the small sample, we tested for equivalence in baseline PPVT scores, receptive and expressive vocabulary performance, self-regulation, work-related skills (see Tables 1, 2, and 3 for descriptive statistics by phase), and demographic factors (gender, age, maternal education, language spoken at home). We found no differences at baseline associated with condition (online supplement Table 2).

To test for effects of intervention condition, we used multiple-level modeling that accounted for nesting of children within classrooms, and we had covariates for site, pretest, age, gender, English language learner (ELL) status, phase, and theme. Comparing pretest to posttest growth in taught words compared with control words for both measures, we found no difference between conditions (see online supplement Table 3). However, according to the Institute of Education Sciences What Works Clearinghouse guideline (2014), the magnitude of the effects was practically important for educational settings (receptive, $d = 0.21$; expressive, $d = 0.20$) and favored the R + P condition over the RO condition. Even though condition effects were not moderated by theme or phase, the benefit of R + P was larger in the spring than in the fall, in particular for expressive knowledge (fall, $d = 0.13$; spring, $d = 0.25$). We also tested for effects of condition on children's posttest scores on the PPVT, peg-tapping task, and teacher ratings of Work-Related Skills and found no effects for condition.

Long-term retention

To determine whether children retained knowledge of the words we taught, they were assessed on taught words after a 4-month delay (see online supplement Table 4). We

Table 2. Spring phase receptive and expressive vocabulary knowledge means (standard deviations) by condition and level of instruction.

	Read + Play (n = 138)		Read Only (n = 84)	
	Pretest	Posttest	Pretest	Posttest
Receptive Vocabulary				
Taught: Percent Correct	0.40 (0.11)	0.58 (0.16)	0.40 (0.12)	0.54 (0.14)
Control: Percent Correct	0.38 (0.19)	0.40 (0.21)	0.37 (0.18)	0.38 (0.20)
Expressive Vocabulary				
Taught: Average Points per Word	0.16 (0.12)	0.52 (0.39)	0.18 (0.17)	0.42 (0.30)
Control: Average Points per Word	0.10 (0.13)	0.12 (021)	0.12 (0.19)	0.11 (0.15)

Note. Pooled means and standard deviations across 10 imputed data sets.

Table 3. Peabody Picture Vocabulary Test and self-regulation means (standard deviations) by condition.

	Read + Play (n = 136)			Read Only (n = 81)		
	Pretest	Posttest	Cohen's d	Pretest	Posttest	Cohen's d
Peabody Picture Vocabulary Test	89.49 (15.23)	90.16 (17.25)	0.04	88.48 (13.44)	91.64 (13.99)	0.20*
Peg Tapping (out of 16)	6.81 (5.68)	8.11 (5.57)	0.24**	6.35 (6.02)	8.56 (5.72)	0.37**
Work-Related Skills (out of 7)	4.48 (1.08)	4.81 (1.13)	0.29**	4.40 (1.19)	4.89 (1.02)	0.44**

Note. Peabody Picture Vocabulary Test (PPVT) standard scores and peg-tapping pretests for children who consented in the fall were completed in the fall. The PPVT and peg-tapping pretests for children who consented in the spring were completed in the spring. Teacher ratings of work-related skills were only available for 124 read + play children and 80 read-only children. Positive values indicate effects of positive pretest to posttest gains. ** $p < .01$.

found strong evidence of learning loss ($p < .01$), with negative effects for receptive vocabulary for RO ($d = -1.17$) and R + P ($d = -1.47$) and expressive vocabulary for RO ($d = -0.79$) and R + P ($d = -0.78$). There was no change in the control words. Tests of condition differences indicated that the loss did not differ by condition, and the two conditions were not significantly different at the delay testing. The loss was less for expressive knowledge, and children still had significantly ($p < .01$) higher expressive knowledge of taught words at the delay testing compared with their expressive knowledge of control words across both conditions, RO ($d = 1.08$) and R + P ($d = 0.74$). At delay, differences in receptive knowledge between taught and control words were not significant.

Fidelity of implementation

The mean fidelity of implementation of book reading was strong for vocabulary picture card review behaviors before reading: 77% in the fall and 87% in the spring. Mean fidelity for behaviors during the reading was marginal: 63% in the fall and 62% in the spring, with scores being depressed because of weak adherence to the strategy of asking children to use the words and comprehension-questioning strategies. We suspect that requesting children to use the words disrupted and slowed the pace of the activity. Comprehension questions may have been neglected for the same reason.

There was much variability in the degree of fidelity to the play methods. In the preplay word review on Days 1 and 2, most teachers had a mean proportion of more than 80% (and all had more than 60%). During the play, there was more variability within classrooms (across the four play videos) and across classrooms both for adherence to vocabulary review behaviors and for adherence to the guided play style. Teachers' average proportion of vocabulary review behaviors during play ranged from 32% to 88%.

Table 4. Tests of pretest to posttest gains by condition.

Construct	B	SE	Cohen's D
Read + Play			
Receptive Vocabulary			
Taught: Percent Correct	0.18	0.01	1.32**
Control: Percent Correct	0.02	0.02	0.11
Expressive Vocabulary			
Taught: Average Points per Word	0.32	0.02	1.12**
Control: Average Points per Word	0.02	0.02	0.10
Taught: Percent Basic or Full Content	0.19	0.01	1.09**
Control: Percent Basic or Full Content	0.02	0.02	0.10
PPVT Standard Scores	0.60	1.26	0.04
Peg Tapping	1.35	0.49	0.24**
Work-Related Skills	0.32	0.10	0.29**
Read Only			
Receptive Vocabulary			
Taught: Percent Correct	0.13	0.01	1.02**
Control: Percent Correct	0.01	0.02	0.04
Expressive Vocabulary			
Taught: Average Points per Word	0.23	0.02	0.94**
Control: Average Points per Word	0.01	0.02	0.01
Taught: Percent Basic or Full Content	0.15	0.02	0.85**
Control: Percent Basic or Full Content	0.02	0.02	0.12
PPVT Standard Scores	2.81	1.08	0.20**
Peg Tapping	2.15	0.61	0.37**
Work-Related Skills	0.49	0.13	0.44**

Note. PPVT = Peabody Picture Vocabulary Test. Positive values indicate effects of positive pretest to posttest gains.
  ** $p < .01$.

Similarly, the average proportion of 1-min segments in which they built on children's contributions (a main indicator of the play style) ranged from 40% to 78%. When we examined the relationship between individual teachers' scores on the two measures of fidelity during play—vocabulary review and play style—there was a significant negative correlation ($r = -.64$, $p < .05$), indicating that teachers who tended to do better on one aspect of fidelity tended to do worse on the other. This finding also meant that no teacher demonstrated high fidelity on both the vocabulary behaviors and the play style during the coded play sessions.

Discussion

We engaged in an extended and iterative process that led to the creation of our intervention methods. Our methods evolved along with our ability to gain insight into word-learning processes in classrooms as we confronted pragmatic challenges such as limitations in instructional time and the need to maintain children's engagement and ensure fidelity of instruction. We created methods for teaching words using book reading and play that were shown to be effective in early stages of the project. In this final year, we sought to determine their effectiveness when used by teachers and to determine if augmenting instruction delivered in group book reading with teacher-directed play resulted in improved learning. Our within-subject analyses of gains in receptive and expressive knowledge provided evidence of strong short-term effects on receptive and expressive knowledge of taught words for both approaches. Between-subjects analyses revealed no evidence of significant differences between the two conditions. Additionally, when words were tested 4 months after instruction was

completed, knowledge had returned roughly to pretest levels. On a more positive note, we found small but positive effects for RO on PPVT scores and evidence of small to moderate gains in self-regulation for both approaches. Fidelity of implementation was adequate for book reading, with the before-book picture card review being excellent, the instruction specific to vocabulary during reading being acceptable, but the comprehension supports being weaker. For play, fidelity to the approach was often weak and variable across different play support strategies.

Learning generated from conducting research in classrooms and with practitioners

It was the intent of the RPL Project to create a scalable teacher-delivered intervention to support young children's vocabulary and self-regulation via evidence-based book reading and guided play strategies. Drawing on the insights of our advisory team, instructional coaches, and teachers, we addressed challenges faced by those who seek to introduce new methods in classrooms. In several cases, those adjustments led us to ask questions and employ strategies that shed new light on learning processes. The pressure to teach as many words as possible in an efficient manner led us to a design in which we taught 16 words with each book, with intensive instruction being delivered for each word on 2 days. The success of that approach lends support to the notion that instruction may enable children to acquire at least initial representations of a significant number of words relatively quickly.

As part of our effort to teach as many words as efficiently as possible, we drew from previous research and theory as we tested the effectiveness of providing different kinds of information about word meanings. When we found all word meanings were roughly equally helpful, we combined them all in our final method. Those definitions contributed to our strong effects on word learning.

Part of our success in teaching so many words might have partly been a byproduct of our use of picture cards associated with gestures. We introduced these cards and gestures to ensure equal instruction and as a means of increasing consistency and fidelity of instruction. The high level of fidelity of use of this method across book reading and play attests to its pragmatic value. Our success may also point to the value of multimodal semantic representations. Words were associated with verbal definitions, story events, pictures from the book, pictures on picture cards, and gestures. Words used in play were further reinforced with the recall of toy figures and play scenarios. Our results lend some support to the added value of the play; further cognitive science work could investigate contributions of visual and gestural knowledge to word learning.

We also gained insight into word learning and the retention of word knowledge when we examined a question of prime importance to educators: Do children remember what they have learned? Our disappointing finding that children returned essentially to baseline levels of performance after a 4-month period during which they likely had almost no exposure to our low-frequency words raises important questions for educators and psychologists. This finding challenges educators to identify strategies for ensuring knowledge is retained and challenges cognitive psychologists to consider what aspects of word knowledge are retained, for how long, what cues can reactivate knowledge previously stored, and how subsequent retrieval might affect previously learned knowledge.

Our efforts to help teachers learn our methods also raise questions about how adults learn new skills that are of interest to educators and cognitive psychologists. During the 12 weeks of intervention, teachers were videotaped and coded for fidelity of implementation, and we found considerable variation in implementation, especially noteworthy for play. In some cases, we suspect that teachers had rarely engaged in such play with children and some may have viewed it as counter to their role as a teacher. We suspect that full understanding of the impediments to use our methods would require exploration of the effects of task demand as they interact with teachers' previous experiences and personal beliefs about the roles of teachers. It is noteworthy that teachers found that viewing video clips of their own teaching and those of others helped them improve. This finding suggests that acquisition of new complex skills is fostered by seeing the full activity enacted. The need for such guidance may be especially great when the activity is novel or is in some way counter to one's beliefs or previous experiences.

Finally, it should be noted that our work proceeded from tightly controlled methods delivered by our trained ISs to development of an approach that used well-structured materials and targeted coaching to help teachers adopt approaches comparable to those we had found to be effective. This sequence of development helped us address the tension between use of tightly controlled methods and implementation in classrooms. Instructional methods must embody evidence-based research, but when they are unduly rigorous or inflexible, teachers cannot teach. The cycles of reflection and revision with teachers and ISs in the current project assisted teachers in the uptake and adoption of methods and fostered a project climate that began to show us how to balance rigor and realism (Snow, 2015).

Learning of taught vocabulary

Both conditions resulted in large effects on receptive and expressive knowledge based on comparisons of pretest to posttest gains on taught versus control words. Book reading followed by play resulted in large effects on the expressive task that indicated average growth of more than one unit of information per word (d = 1.12). Expressive gains in the RO condition were also substantial (RO, d = 0.92). Large effects on receptive knowledge of taught words were found for both conditions (R + P, d = 1.32; RO, d = 1.02). There were no statistically significant differences between conditions in effects on the learning of target words.

Our results for receptive knowledge gains were more than double the average size found in previous vocabulary intervention studies. Mol and Bus (2011) conducted a meta-analysis of 31 book-reading studies and found a fixed effect of d = 0.45, and the National Early Literacy Panel (2009) found average effects of d = 0.45 when an outlier was excluded. Growth in expressive knowledge in our data was substantially greater than the effect of d = 0.66 found by Mol, Bus and de Jong (2009) and expressive gains found by Neuman and Dwyer (2011; d = 0.64), who used book reading, supplementary activities, and videos. The current study's gains in expressive knowledge also were stronger than the effects in pretest to posttest knowledge of d = 0.70 found by McKeown and Beck (2014). They taught 10 words over four readings. The strong gains in expressive knowledge, our measure of depth of learning, are encouraging given the association between depth measures and reading comprehension (National Early Literacy Panel, 2009; Proctor et al., 2012).

Why might our approach have been so effective? It offers highly concentrated exposure to multiple forms of information about word meanings at the point when words are heard, and we shifted responsibility from the teacher to the child across instructional occasions. Our prompts shifted from requesting recognition responses, to producing words when shown a picture, to more independent responding. Children also were encouraged to fill in gaps in a sentence in which the target word was used, thereby using syntactic and semantic cues. Children were challenged to draw on semantic knowledge as they produced words after hearing definitions and to give definitions when asked. Pictures, a commonly recognized strategy for use with ELLs (Pollard-Durodola et al., 2011), provided visual support. Pictures also were helpful to teachers. They ensured that each word was discussed and provided visual supports for comments about word meanings. The high fidelity of implementation score for the picture card reviews attests to their use. Focus-group discussions reinforced this point, because teachers were unanimous in their positive feelings about pictures. Illustrations in the books also provided a way to anchor the word to story events. Also, words were heard in stories that held children's attention and were read and discussed four times. They supplied a context for associating words with a network of terms (e.g., ladle, soup, stir, cook). Word learning could have been enhanced by these efforts to build understanding, because instruction designed to foster understanding of story events has been found to offer nearly as much support for word learning as verbal definitions (Dickinson et al., in press).

After the 4-month delay, our receptive measure failed to reveal sustained learning, but our expressive task continued to reveal large effects across both conditions, RO ($d = 1.08$) and R + P ($d = 0.74$). This evidence of sustained learning is consistent with earlier work on fast mapping that has shown retention over several days (Dickinson, 1984) and weeks (Carey, 1978). The fact that there was significant learning loss after a 4-month delay was not surprising given that our words likely were not encountered in everyday conversations or in other books. Future research should examine whether brief reviews of previously taught words help maintain previously taught vocabulary.

Condition effects

Taught words

We hypothesized that the R + P condition would result in greater learning. We failed to find statistical evidence of a benefit from this condition, a result that might be partly accounted for by our small sample. However, there was a larger effect for R + P than for RO (i.e., differences of $d = 0.30$ on receptive; $d = 0.20$ on expressive). It is tempting to suggest play was the active ingredient that resulted in that stronger effect, but we cannot draw that conclusion definitively because unlike the R + P condition, the RO condition did not involve formal exposure to words outside the book readings. Data from previous phases of this project showed, however, that children learned words better when they were taught through book reading and play than when they were taught through book reading and a picture card vocabulary review activity, similar to the RO condition (Dickinson et al., in press). Those findings suggest that the play in our current study might contribute more to vocabulary gains than mere additional exposure.

One reason for the lack of a significant condition effect may reside with teachers' fidelity to the play protocols. When the hypothesized effects of an intervention are not

observed, it is important to examine the degree to which the intervention was implemented as designed (Mendive, Weiland, Yoshikawa, & Snow, 2016). Although teachers' fidelity to vocabulary review prior to play was reasonably high, there was much variability in teachers' adherence to both the vocabulary review and the desired play style during the play. In fact, there was a significantly negative correlation between teachers' use of desired vocabulary review behaviors and the guided-play style during play sessions. If a teacher was more adept at play, then she was not adequately reviewing the words, and vice versa. Therefore, we do not know if the implementation of the guided-play sessions in the manner we envisioned would be an effective addition to the vocabulary support during book reading. Evidence supporting that suggestion came from results from previous phases of the project (Dickinson et al., in press). Those earlier data showed that reviewing words taught through book reading in the context of adult-supported play was more effective than having children simply play freely after the book-reading sessions. There may be a threshold level of fidelity required for play to have an impact on word learning, and our teachers were unable to reach that level under the circumstances of this study.

Many factors could account for why our teachers struggled to achieve fidelity to the method. The training period was brief, and subsequent coaching interactions were limited in duration due to demands on teachers' time. Also, teachers had little preparation time during the school day and often carried heavy responsibilities at home, precluding additional time for preparation beyond school hours. Our ISs' observations and conversations with teachers led them to speculate that teachers spent minimal time preparing for their reading and play sessions. Guided play was especially challenging because of its novelty. Teachers are familiar with book reading, but the play approach was foreign to them and some were uncomfortable engaging in fantasy play. This discomfort could be the result of a lack of experience or a conflict with their view of what it means to be a teacher. Also, our approach to play may have conflicted with some teachers' views of their role in children's play. Free play is common in early childhood classrooms, but many teachers do not actively engage in it, partially because of a concern that they would interfere with children's play (Pyle & Danniels, 2017). Researchers have reported that even when teachers do participate, they sometimes struggle with juggling playfulness and language support strategies, such as open-ended questions (Meacham, Vukelich, Han, & Buell, 2014). It may be that learning to follow children's lead in play and employing the multiple language support methods involved in our protocol require more investments of teachers' time and/or more coaching than we provided.

Our data suggest that play, as it was implemented, did not make a substantial contribution to word learning. However, it could be the case that there is a limit to how much children can learn about new words in a short time. Maybe if we had compared play alone with play plus reading we would have found a similar result, with play accounting for substantial word learning and book reading supplying only a small added boost. We currently are investigating the amount of learning that may occur through playful learning apart from book reading.

Generalized learning

We found significant gains on age-normed scores on the PPVT for the RO condition ($d = 0.20$) but not for R + P. The pretest to posttest gains on the PPVT for RO were statistically

significant, and the small effect ($d = 0.20$) was educationally significant. Because our taught words were unusual, they would not have been among those sampled by the PPVT. Thus, gains cannot be attributed to direct instruction. What might account for why we found generalized gains for only the RO condition? One possible explanation is that teachers might have become more attuned to the need to provide direct instruction on word meanings, more facile at engaging in conversations about words, and more likely to sprinkle their talk with varied and less common words. These improved language instruction methods might have been used during incidental opportunistic instruction during times when children engage in independent activities. Reading-only teachers had more time than R + P teachers for the type of opportunistic interactions they were accustomed to using while children are engaged in activities of their own choosing. In contrast, R + P were involved in small-group guided play, a setting that was new to them. Our fidelity data indicated that many struggled to use the play as effectively as we hoped as a platform for the type of responsive interactions known to be supportive of language. Support for the value of the type of incidental conversations during free play in which RO teachers had more time to engage came from the finding that teacher–child talk during free play has been found to be associated with vocabulary learning (Dickinson & Porche, 2011). Encouragement for such interactions is part of successful language-focused interventions (Wasik et al., 2006; Wilson et al., 2013).

Both conditions were associated with significant growth in peg tapping and Work-Related Skill measures of self-regulation (R + P, $d = 0.24$ and 0.29, respectively; RO, $d = 0.37$ and 0.44). However, this measure is not age-adjusted, so considerable caution is required when interpreting this outcome. This finding lends some support to our hypothesis that children may gain some capacity to self-regulate as a result of participating in language-rich classroom activities. Our approach might have fostered improved self-regulation because gains in language are associated with improvements in self-regulation (Bohlmann et al., 2015; Nesbitt et al., 2015). A second possibility is that by engaging in group activities such as book reading and play, children learn to be more attentive. However, we must be cautious in drawing conclusions: Our measures were not normed, and previous observational studies have also found age-related growth on these measures (Lipsey et al., 2017).

Limitations

Our method relied on contrasting the effects of two intervention conditions; we did not have a control group. The lack of control group meant we relied on children to serve as their own controls for tests of taught words versus nontaught vocabulary. Although this approach allowed for strong control of individual variability when assessing vocabulary learning, we could not determine if gains in self-regulation simply reflected typical developmental improvement. Moreover, target vocabulary words were sometimes difficult to integrate into play, unlikely to be encountered or used at other times, and unlikely to boost scores on the PPVT. Our play might have resulted in greater gains if the words were easier for both teachers and children to incorporate into play. Given our small sample size, the results may have been affected by the ability of particular teachers to implement our methods. This issue was particularly important for the play approach, because teachers found it to be relatively challenging to implement. We also do not know if teachers generalized their attention to words to other classroom settings and content areas. For

classrooms to have significant effects on children's vocabulary knowledge, attention to words likely needs to occur throughout the day and across time.

Conclusions and future directions

Teachers mastered elements of our book-reading method that related specifically to the teaching of new words, but they demonstrated less facility-supporting discussion designed to foster comprehension. We do not know if teachers can generalize our methods to other books without the support of instructional materials or if they can sustain use of our methods without coaching.

Implementing guided play was challenging. Often teachers found it difficult to find time to schedule children for play in a small group of sufficient length and frequency to enable them to gain familiarity with the story line and proficiency in using the language of the story. Doing so required a capable assistant teacher and commitment by the teacher to allocate needed time. Also, as noted, it was difficult for teachers to engage in guided play using the methods as we designed them. As such, the guided-play condition became more directed than would have been optimal. Additional coaching, possibly with greater use of videos of effective methods, would be helpful. A play-based approach like ours does not have to use such high-level words; we needed them because our test of learning required that children not know the words prior to the study. Using words that are within children's proximal level of development might have enabled those with weaker skills to begin using new words and resulted in evidence of generalized learning. This adjustment might be particularly helpful to English-language learners.

We currently are exploring the possibility that games, which are a different type of play, might be better aligned with classroom life and the need to address specific learning goals. With games, established rules and procedures help to focus children's play by often drawing their attention to particular content or providing opportunities to practice relevant skills, while also enabling them to feel the excitement that comes from chance and participation (Hassinger-Das et al., 2017b). A game that integrates vocabulary review has been found to effectively help support children's word-learning (Hassinger-Das et al., 2016).

While the intervention was designed to meet the needs of both less and more advanced language learners, prior language was still related to word learning. Effects were more evident on the expressive task, our task evaluating depth of knowledge, than our receptive measure. This differential effect deserves future investigation using measures of initial and deeper learning because it provides further evidence that word learning is affected by semantic knowledge.

Given the time constraints in preschool classrooms and the challenges teachers face in choosing play types that match their teaching styles, we anticipate that, as we continue to develop and refine our methods, the movement between applied settings and basic cognitive science will continue to yield insights relevant to teaching and learning and reveal issues worthy of in-depth examination by cognitive scientists.

Acknowledgments

Funding

References

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

Berk, L. E., & Spuhl, S. T. (1995). Maternal interaction, private speech, and task-performance in preschool-children. *Early Childhood Research Quarterly*, *10*(2), 145–169. doi:10.1016/0885-2006(95)90001-2

Biemiller, A. (2010). *Words worth teaching: Closing the vocabulary gap*. Columbus, OH: McGraw Hill.

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

Bindman, S. W., Hindman, A. H., Bowles, R. P., & Morrison, F. J. (2013). The contributions of parental management language to executive function in preschool children. *Early Childhood Research Quarterly*, *28*(3), 529–539. doi:10.1016/j.ecresq.2013.03.003

Bivens, J. A., & Berk, L. E. (1990). A longitudinal study of the development of elementary school children's private speech. *Merrill-Palmer Quarterly-Journal of Developmental Psychology*, *36*(4), 443–463.

Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology*, *101*(2), 294–304. doi:10.1037/a0013844

Bohlmann, N. L., Maier, M. F., & Palacios, N. (2015). Bidirectionality in self-regulation and expressive vocabulary: Comparisons between monolingual and dual language learners in preschool. *Child Development*, *86*(4), 1094–1111. doi:10.1111/cdev.12375

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities*, *44*(5), 431–443. doi:10.1177/0022219411410042

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology*, *40*(6), 1105–1122. doi:10.1037/0012-1649.40.6.1105

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Brookline, MA: Brookline Books.

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the early childhood longitudinal study (ecls) kindergarten to first grade sample. *Journal of Educational Psychology*, *98*(3), 489–507. doi:10.1037/0022-0663.98.3.489

Cooper, D. H., & Farran, D. C. (1988). Behavioral risk factors in kindergarten. *Early Childhood Research Quarterly*, *3*(1), 1–19. doi:10.1016/0885-2006(88)90026-9

Coyne, M. D., McCoach, D. B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly*, *30*(2), 74–88. doi:10.2307/30035543

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *Elementary School Journal*, *110*(1), 1–18. doi:10.1086/598840

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(6), 934–945. doi:http://dx.doi.org/10.1037/0012-1649.33.6.934

Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say, not as I do". *Developmental Psychobiology*, *29*(4), 315–334. doi:10.1002/(sici)1098-2302(199605)29:4<315::aid-dev2>3.3.co;2-c

Dickinson, D. K. (1984). First impressions: Children's knowledge of words after a single exposure. *Journal of Applied Psycholinguistics*, *5*, 359–373. doi:10.1017/s0142716400005233

Dickinson, D. K. (2011). Teachers' language practices and academic outcomes of preschool children. *Science*, *333*(6045), 964–967. doi:10.1126/science.1204526

Dickinson, D. K., Freiberg, J. B., & Barnes, E. M. (2011). Why are so few interventions really effective?: A call for fine-grained research methodology. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (Vol. 3, pp. 337–357). New York, NY: Guilford.

Dickinson, D. K., Hofer, K. G., Barnes, E. M., & Grifenhagen, J. B. (2014). Examining teachers' language in head start classrooms from a systemic linguistics approach. *Early Childhood Research Quarterly*, *29*, 231–244. doi:10.1016/j.ecresq.2014.02.006

Dickinson, D. K., Nesbitt, K.T., Collins, M.F., Hadley, E.,B., Newman, K.,, Rivera, B.L, . . . & Hirsh-Pasek, K. (in press). Teaching for breadth and depth of vocabulary knowledge: Learning from explicit and implicit instruction and the storybook texts. Early Childhood Research Quarterly.

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, *82*(3), 870–886. doi:10.1111/j.1467-8624.2011.01576.x

Dickinson, D. K., & Tabors, P. O. (Eds.). (2001). *Beginning literacy with language: Young children learning at home and school.* Baltimore, MD: Brookes Publishing.

Dionne, G., Tremblay, R., Boivin, M., Laplante, D., & Perusse, D. (2003). Physical aggression and expressive vocabulary in 19-month-old twins. *Developmental Psychology*, *39*(2), 261–273. doi:10.1037//0012-1649.39.2.261

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. doi:10.1037/0012-1649.43.6.1428

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test, fourth edition*. Minneapolis, MN: NCS Pearson.

Fernyhough, C., & Fradley, E. (2005). Private speech on an executive task: Relations with task difficulty and task performance. *Cognitive Development*, *20*(1), 103–120. doi:10.1016/j.cogdev.2004.11.002

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. B. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology*, *50*(6), 1698–1709. doi:10.1037/a0036633

Fisher, K., Hirsh-Pasek, K., Golinkoff, R. M., Singer, D. G., & Berk, L. (2011). Playing around in school: Implications for learning and educational policy. In A. D. Pellegrini (Ed.), *Oxford library of psychology. The Oxford handbook of the development of play* (pp.341–360). New York, NY: Oxford University Press.

Gonzalez, J. E., Pollard-Durodola, S., Simmons, D. C., Taylor, A. B., Davis, M. J., Kim, M., & Simmons, L. (2010). Developing low-income preschoolers' social studies and science vocabulary knowledge through content-focused shared book reading. *Journal of Research on Educational Effectiveness*, *4*(1), 25–52. doi:10.1080/19345747.2010.487927

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323. doi:10.1037/1082-989x.11.4.323

Hadley, E. B. (2017). Understanding, measuring, and fostering preschool children's acquisition of vocabulary depth. (PhD), Vanderbilt University, Nashville, TN.

Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., Golinkoff, R. M., & Nesbitt, K. T. (2016). Examining the acquisition of vocabulary knowledge depth among preschool students. *Reading Research Quarterly*, *51*(2), 181–198. doi:10.1002/rrq.130

Hammond, S. I., Muller, U., Carpendale, J. I. M., Bibok, M. B., & Liebermann-Finestone, D. P. (2012). The effects of parental scaffolding on preschoolers' executive function. *Developmental Psychology*, *48*(1), 271–281. doi:10.1037/a0025519

Han, M., Moore, N., Vukelich, C., & Buell, M. (2010). Does play make a difference? Effects of play intervention on at-risk preschoolers' vocabulary learning. *American Journal of Play*, *3*, 82–105.

Harris, J., Golinkoff, R. M., & Hirsh-Pasek, K. (2011). Lessons from the crib for the classroom: How children really learn vocabulary. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 49–66). New York, NY: Guilford.

Hassinger-Das, B., Ridge, K., Parker, A., Golinkoff, R. M., Hirsh-Pasek, K., & Dickinson, D. K. (2016). Building vocabulary knowledge in preschoolers through shared book-reading and game-play. *Mind, Brain, and Education*, *10*, 71–80. doi:10.1111/mbe.12103

Hassinger-Das, B., Toub, T. S., Hirsh-Pasek, K., & Golinkoff, R. M. (2017a). A matter of principle: Applying language science to the classroom and beyond. *Translational Issues in Psychological Science*, *3*, 5–18. doi:10.1037/tps0000085

Hassinger-Das, B., Toub, T. S., Zosh, J. M., Michnick, J., Hirsh-Pasek, K., & Golinkoff, R. M. (2017b). More than just fun: A place for games in playful learning. *Infancia Y Aprendizaje*, *40*, 191–218. doi:10.1080/02103702.2017.1292684

Hindman, A. H., Wasik, B. A., & Erhart, A. C. (2012). Shared book reading and Head Start preschoolers' vocabulary learning: The role of book-related discussion and curricular connections. *Early Education & Development*, *23*(4), 451–474. doi:10.1080/10409289.2010.537250

Hirsh-Pasek, K., & Golinkoff, R. (2012). How babies talk: Six principles of early language development. In S. Odom, E. Pungello, & Gardner-Neblett (Eds.), *Infants, toddlers, and families in poverty: Research implications for early childcare* (pp. 77–100). New York, NY: Guilford.

Hirsh-Pasek, K., Golinkoff, R. M., Berk, L., & Singer, D. (2009). *A mandate for playful learning in preschool: Presenting the evidence*. New York, NY: Oxford University Press.

IBM Corp. Released. (2013). *IBM SPSS statistics for windows, version 22.0*. Armonk, NY: Author.

Justice, L. M., Mashburn, A., Pence, K. L., & Wiggins, A. (2008). Experimental evaluation of a preschool language curriculum: Influence on children's expressive language skills. *Journal of Speech Language and Hearing Research*, *51*(4), 983–1001. doi:10.1044/1092-4388(2008/072)

Justice, L. M., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech & Hearing Services in Schools*, *36*(1), 17–32. doi:10.1044/0161-1461(2005/003)

Landry, S. H., Miller-Loncar, C. L., Smith, K. E., & Swank, P. R. (2002). The role of early parenting in children's development of executive processes. *Developmental Neuropsychology*, *21*(1), 15–41. doi:10.1207/s15326942dn2101_2

Landry, S. H., Smith, K. E., Swank, P. R., & Miller-Loncar, C. L. (2000). Early maternal and child influences on children's later independent cognitive and social functioning. *Child Development*, *71*(2), 358–375. doi:10.1111/1467-8624.00150

Leider, C. M., Proctor, C. P., Silverman, R. D., & Harring, J. R. (2013). Examining the role of vocabulary depth, cross-linguistic transfer, and types of reading measures on the reading comprehension of latino bilinguals in elementary school. *Reading and Writing*, *26*(9), 1459–1485. doi:10.1007/s11145-013-9427-6

Lillard, A. (2016). *Montessori: The science behind the genius*. New York, NY: Oxford University Press.

Lipsey, M. W., Nesbitt, K. T., Farran, D. F., Dong, N., Fuhs, M. W., & Wilson, S. J. (2017). Cognitive self-regulation measures for prekindergarten children that perform well for predicting academic achievement: A comparative evaluation. *Journal of Educational Psychology*, *109*(8), 1084–1102. doi:10.1037/edu0000203

Loftus, S. M., Coyne, M. D., McCoach, D. B., Zipoli, R., & Pullen, P. C. (2010). Effects of a supplemental vocabulary intervention on the word knowledge of kindergarten students at risk for language and literacy difficulties. *Learning Disabilities Research & Practice*, *25*(3), 124–136. doi:10.1111/j.1540-5826.2010.00310.x

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research*, *80*(3), 300–335. doi:10.3102/0034654310377087

McKeown, M. G., & Beck, I. L. (2014). Effects of vocabulary instruction on language processing: Comparing two approaches. *Early Childhood Research Quarterly*, 29, 520–530. doi:10.1016/j.ecresq.2014.06.002

Meacham, S., Vukelich, C., Han, M., & Buell, M. (2014). Preschool teachers' questioning in socio-dramatic play. *Early Childhood Research Quarterly*, 29(4), 562–573. doi:10.1016/j.ecresq.2014.07.001

Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Educational Psychology*, 108(1), 130–145. doi:10.1037/edu0000047

Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, 137(2), 267–296. doi:10.1037/a0021890

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, 79(2), 979–1007. doi:10.3102/0034654309332561

National Center for Educational Sciences. (2013). *The nation's report card: Trends in academic progress, 2012*. Washington D.C.: Retrieved from http://nces.ed.gov/nationsreportcard/pubs/main2012/2013456.aspx.

National Early Literacy Panel. (2009). Developing early literacy: Report of the national early literacy panel. Retrieved from http://lincs.ed.gov/publications/pdf/NELPReport09.pdf

Nesbitt, K. T., Farran, D. C., & Fuhs, M. W. (2015). Executive function skills and academic achievement gains in prekindergarten: Contributions of learning-related behaviors. *Developmental Psychology*, 51(7), 865–878. doi:10.1037/dev0000021

Neuman, S. B., & Dwyer, J. (2011). Developing vocabulary and conceptual knowledge for low-income preschoolers: A design experiment. *Journal of Literacy Research*, 43(2), 103–129. doi:10.1177/1086296x11403089

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational effects of a vocabulary intervention on preschoolers' word knowledge and conceptual development: A cluster-randomized trial. *Reading Research Quarterly*, 46(3), 249–272. doi:10.1177/1086296x11403089

NICHD ECCRN. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology*, 41(2), 428–442. doi:10.1037/0012-1649.41.2.428

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology*, 94(1), 23–33. doi:10.1037/0022-0663.94.1.23

Pollard-Durodola, S. D., Gonzalez, J. E., Simmons, D. C., Kwok, O., Taylor, A. B., Davis, M. J., … Simmons, L. (2011). The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. *Exceptional Children*, 77(2), 161–183. doi:10.1177/001440291107700202

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on head start teachers and children. *Journal of Educational Psychology*, 102(2), 299–312. doi:10.1037/a0017763

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing*, 25(7), 1635–1664. doi:10.1007/s11145-011-9336-5

Pullen, P. C., Tuckwiller, E. D., Konold, T. R., Maynard, K. L., & Coyne, M. D. (2010). A tiered intervention model for early vocabulary instruction: The effects of tiered instruction for young students at risk for reading disability. *Learning Disabilities Research & Practice*, 25(3), 110–123. doi:10.1111/j.1540-5826.2010.00309.x

Pyle, A., & Danniels, E. (2017). A continuum of play-based learning: The role of the teacher in play-based pedagogy and the fear of hijacking play. *Early Education and Development*, 28(3), 274–289. doi:10.1080/10409289.2016.1220771

Roskos, K., & Burstein, K. (2011). Assessment of the design efficacy of a preschool vocabulary instruction technique. *Journal of Research in Childhood Education*, 25, 268–287. doi:10.1080/02568543.2011.580041

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: John Wiley & Sons, Inc.

Silverman, R., & Crandell, J. D. (2010). Vocabulary practices in prekindergarten and kindergarten classrooms. *Reading Research Quarterly*, *45*(3), 318–340. doi:10.1598/rrq.45.3.3

Smith, L. E., Borkowski, J. G., & Whitman, T. L. (2008). From reading readiness to reading competence: The role of self-regulation in at-risk children. *Scientific Studies of Reading*, *12*(2), 131–152. doi:10.1080/10988430801917167

Snow, C. E. (2015). Rigor and realsim. Doing educational science in the real world. *Educational Researcher*, *44*(9), 460–466. doi:10.3102/0013189X15619166

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–401. doi:10.1598/rrq.21.4.1

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*, 934–947. doi:10.1037/0012-1649.38.6.934

Toub, T. S., Hassinger-Das, B., Nesbitt, K.T., Ilgaz, H., Weisberg, D. S., Hirsh-Pasek, K., … & Dickinson, D.. (2018). The language of play: Developing preschool vocabulary through play following shared book-reading. *Early Childhood Research Quarterly*, *45*, 1–7.

Vygotsky, L. (1978). *Mind in society: The formation of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wasik, B. A., Bond, M. A., & Hindman, A. (2006). The effects of a language and literacy intervention on head start children and teachers. *Journal of Educational Psychology*, *98*(1), 63–74. doi:10.1037/0022-0663.98.1.63

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130. doi:10.1111/cdev.12099

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Kittredge, A. K., & Klahr, D. (2016). Guided play: Principles and practices. *Current Directions in Psychological Science*, *25*(3), 177–182. doi:10.1177/0963721416645512

Weisberg, D. S., Ilgaz, H., Hirsh-Pasek, K., Golinkoff, R., Nicolopoulou, A., & Dickinson, D. K. (2015). Shovels and swords: How realistic and fantastical themes affect children's word learning. *Cognitive Development*, *35*, 1–14. doi:10.1016/j.cogdev.2014.11.001

What Works Clearinghouse (2014). Procedures and Standards Handbook (Version 3.0). Retrieved from https://ies.ed.gov/ncee/wwc/Handbooks/

Wilson, S. J., Dickinson, D. K., & Rowe, D. W. (2013). Impact of an early reading first program on the language and literacy achievement of children from diverse language backgrounds. *Early Childhood Research Quarterly*, *28*(3), 578–592. doi:10.1016/j.ecresq.2013.03.006