The language of play: Developing preschool vocabulary through play following shared book-reading

Check for updates

Tamara Spiewak Toub [a], Brenna Hassinger-Das [a,*], Kimberly Turner Nesbitt [b,1], Hande Ilgaz [a,2], Deena Skolnick Weisberg [a,3], Kathy Hirsh-Pasek [a], Roberta Michnick Golinkoff [c], Ageliki Nicolopoulou [d], David K. Dickinson [e]

[a] Department of Psychology, Temple University, 1701 North 13th Street, 6th Floor Weiss Hall, Philadelphia, PA 19122, United States
[b] Department of Teaching and Learning, Peabody College, Vanderbilt University, 230 Peabody, 230 Appleton Place, Nashville, TN 37203, United States
[c] School of Education, University of Delaware, 206 Willard Hall, Newark, DE 19716, United States
[d] Department of Psychology, Lehigh University, 17 Memorial Drive East, Chandler-Ullman Hall, Bethlehem, PA 18015, United States
[e] Peabody Research Institute, Vanderbilt University, 230 Appleton Place, PMB 181, Nashville, TN 37203, United States

A B S T R A C T

Two studies explored the role of play in a vocabulary intervention for low-income preschoolers. Both studies presented new vocabulary through book-readings. Study 1 children ($N = 249$; $M_{age} = 59.19$ months) were also randomly assigned to participate in Free Play, Guided Play, or Directed Play with toys relating to the books. Guided and Directed Play conditions involved different styles of adult support. Although children in all conditions showed significant gains in knowledge of target vocabulary words, children in both adult-supported conditions showed significantly greater gains than children experiencing Free Play. In Study 2, classroom teachers implemented our procedures instead of researchers. All children ($N = 101$; $M_{age} = 58.65$ months) reviewed half the vocabulary words through a hybrid of guided and directed play and half the words through a picture card review activity. Children showed significant pre- to post-test gains on receptive and expressive knowledge for both sets of taught words, but they also showed significantly greater expressive vocabulary gains for words reviewed through play. These results suggest that there are unique benefits of adult-supported play-based activities for early vocabulary growth.

© 2018 Elsevier Inc. All rights reserved.

One proverb states, "All work and no play makes Jack a dull boy" (Howell, 1659). Yet, "All play and no work makes Jack a mere toy" (Edgeworth, 1825, p. 155). This poetic pair captures not only the benefits of children's play but also its limitations. Time for play in classrooms is dwindling, as educators and policymakers emphasize other activities believed to be more educationally effective (Elkind, 2008; Miller & Almon, 2009). However, data show

that playful learning promotes language, cognitive, and social skills that are critical for academic success (e.g., Hirsh-Pasek, Golinkoff, Berk, & Singer, 2009; Roskos & Christie, 2000). Although a recent review of the research on the developmental benefits of play (Lillard et al., 2013) provided an important critique, the concerns are not equally applicable to all types of play or all developmental outcomes (Nicolopoulou & Ilgaz, 2013). To balance conflicting notions of the role of play in children's development and education, we need to understand which types of play activities relate to which outcomes. As part of the broader Read-Play-Learn project, two studies presented here explored various approaches to leveraging play alongside more established book-reading methods to promote vocabulary growth in low-income preschoolers. Play provides a way to scaffold children's vocabulary development through engagement with words in meaningful contexts. A playful review of new vocabulary items differs substantially from the didactic teaching methods commonly used in preschool classrooms (Early et al., 2010). Here we focus on vocabulary introduced through book-reading sessions and evaluate whether playful approaches activities to word learning assist low-income preschoolers in augmenting their vocabulary knowledge. We also examine whether

* Corresponding author. Present address: 41 Park Row, Room 1324, New York, NY 10038, United States.
E-mail addresses: tamara.spiewak.toub@temple.edu
(T.S. Toub), bhassingerdas@pace.edu (B. Hassinger-Das),
Kimberly.Nesbitt@unh.edu (K.T. Nesbitt), hande.ilgaz@bilkent.edu.tr (H. Ilgaz),
deena.weisberg@psych.upenn.edu (D.S. Weisberg), khirshpa@temple.edu
(K. Hirsh-Pasek), Roberta@udel.edu (R.M. Golinkoff), agn3@lehigh.edu
(A. Nicolopoulou), david.dickinson@vanderbilt.edu (D.K. Dickinson).
[1] Present address: Department of Human Development and Family Studies, University of New Hampshire, 217 Pettee Hall, Durham, NH 03824, United States.
[2] Present address: Department of Psychology, Bilkent University, 06800 Bilkent, Ankara, Turkey.
[3] Present address: Department of Psychology, University of Pennsylvania, 425 S. University Ave., Philadelphia, PA 19104, United States.

play is more effective than a direct instruction approach to supplementing the vocabulary support occurring within book-reading.

1. Promoting vocabulary growth in low-income preschoolers

Early vocabulary powerfully predicts children's later language development, reading skills, school-readiness, and academic success (Dickinson & Porche, 2011; Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015; Rowe, Raudenbush, & Goldin-Meadow, 2012). For example, Storch and Whitehurst (2002) found statistically significant indirect effects of preschool oral language skills (e.g., receptive and expressive vocabulary) on reading abilities in Grades 1–4. Also, rates of early vocabulary growth predict the structure of children's brains years later, with changes in cortical regions relevant to continued language development (Asaridou, Demir-Lira, Goldin-Meadow, & Small, 2017).

Although fostering vocabulary development is important for all children, it is especially crucial to examine trajectories for children from low-income families (Farkas & Beron, 2004; Snow, Burns, & Griffin, 1998). The quantity and quality of language input for children from low-income families is typically less than for their higher-income peers (Hart & Risley, 1995; Hindman, Wasik, & Snell, 2016; Hoff, 2013; Rowe, 2008). Both quantity and quality of early language input play a large role in language development (Cartmill et al., 2013; Goldin-Meadow et al., 2014; Hirsh-Pasek, Adamson et al., 2015), and quality often mediates the relation between socioeconomic status and language outcomes (Bracken & Fischel, 2008). Enhancing language environments in formal childcare or preschool initiatives (e.g., Head Start, universal pre-K) is one potential venue for providing language support to children otherwise at risk (Vernon-Feagans, Bratsch-Hines, & The Family Life Project Key Investigators, 2013), and that is what we aimed to do in the current study.

1.1. Vocabulary development in early childhood programs

Much work needs to be done to ensure that early childhood programs are providing high-quality developmental support for all children. Numerous studies have found low-quality language and literacy instruction in early childhood classrooms (Dickinson, Hofer, Barnes, & Grifenhagen, 2014; Hindman & Wasik, 2013; Justice, Mashburn, Hamre, & Pianta, 2008; Mashburn et al., 2008). Early word learning, in particular, is not sufficiently supported at many preschools (National Early Literacy Panel, 2008; Neuman & Dwyer, 2009). Recent data from Fuller, Bein, Bridges, Kim, and Rabe-Hesketh (2017) indicate that pre-kindergarten programming that is academically oriented (i.e., which spends significant amounts of time emphasizing oral language skills and preliteracy skills, as well as math concepts) is associated with greater gains in children's language and math skills than those seen for children in home-based care or less academic-oriented programming. Such research further suggests that incorporating high-quality support for language development is a key ingredient in maximizing the benefits of early education.

Part of high-quality support for vocabulary growth involves addressing the many aspects of what it means to know a word well. Vocabulary knowledge can be conceptualized as being along a continuum from not understanding a word's meaning to having a deep understanding, and vocabulary interventions and assessments differ in the level of knowledge targeted (Christ & Wang, 2011; Coyne, McCoach, Loftus, Zipoli Jr, & Kapp, 2009; Hadley, Dickinson, Hirsh-Pasek, Golinkoff, & Nesbitt, 2016). Although many efforts to improve children's word knowledge and the word gap focus on presenting many new words efficiently, children show greater depth of vocabulary knowledge when they engage with new words in extended and varied ways (Coyne et al., 2009).

1.2. Efficacy of vocabulary interventions

Research on vocabulary acquisition suggests that word learning is facilitated through six principles: (1) frequent exposure, (2) capturing the child's interest, (3) interactive and responsive environments, (4) meaningful context, (5) diversity of words and language structures, and (6) leveraging of grammatical knowledge (Harris, Golinkoff, & Hirsh-Pasek, 2011; Hassinger-Das, Toub, Hirsh-Pasek, & Golinkoff, 2017; Konishi, Kanero, Freeman, Golinkoff, & Hirsh-Pasek, 2014; Reed, Hirsh-Pasek, & Golinkoff, 2017). Wasik, Hindman, and Snell (2016) particularly emphasize the value of systemic exposures to new words and a variety of tasks that invite children to truly engage with the words.

Our work focuses on common activities that can implement these principles within the preschool classroom: reading and playing. During these activities, vocabulary words and their meanings can be highlighted in interesting and relevant ways for children through explicit verbal discussions and images or props that depict word meanings, as well as gestures that provide nonverbal support (Goldin-Meadow & Alibali, 2013). Reading and playing are promising ways to go beyond rote memorization and facilitate children's deeper and longer-lasting word knowledge. Furthermore, playing especially enables the kind of adaptive and responsive interactions that should facilitate the natural learning of vocabulary.

1.3. Shared book-reading to support vocabulary

An extensive review of preschool and kindergarten interventions that were designed to build language found that shared book-reading is the intervention that most consistently results in vocabulary growth (National Early Literacy Panel, 2008). Meta-analyses that included studies completed after the cut-off point of the National Early Literacy Panel (Marulis & Neuman, 2010; Mol, Bus, de Jong, & Smeets, 2008; Mol, Bus, & Jong, 2009) continue to find consistent evidence that book-reading interventions yield moderate-to-strong effects on preschoolers' vocabulary. Although a recent systematic review of book-reading interventions (Wasik et al., 2016) noted that gains are typically modest in terms of the proportion of words that are learned, the authors also argue that book-reading lends itself to the types of opportunities for exposure to and engagement with words that best facilitate learning. We therefore used book-reading in our program and incorporated evidence-based strategies to enhance vocabulary growth. For example, we built in multiple exposures to words through repeated readings, questioning about words and the story, and provision of explicit definitions (Beck & McKeown, 2007; Biemiller & Boote, 2006; Coyne, McCoach, & Kapp, 2007; Coyne et al., 2009; Wasik, Bond, & Hindman, 2006). We also added opportunities for children to say target words (Sénéchal, 1997) and to observe and use gesture for non-verbal expression of word meaning (Goldin-Meadow & Alibali, 2013; Rowe, Silverman, & Mullan, 2013). More details on our book-reading approach can be found in Dickinson and colleagues (in preparation).

1.4. Play-based activities to further support vocabulary

Play might be another particularly effective component for vocabulary interventions (Weisberg, Zosh, Hirsh-Pasek, & Golinkoff, 2013). Play is often inherently interesting, interactive, and meaningful for children, and the interdisciplinary science of learning literature indicates that learning is maximized through such meaningful and socially interactive learning environments in which children are active and engaged (Chi, 2009; Hirsh-Pasek,

Zosh et al., 2015; Meltzoff, Kuhl, Movellan, & Sejnowski, 2009). Skill-specific curricula often embed focused academic instruction within play, and studies suggest that these curricula are more effective than less targeted approaches in promoting the relevant academic skills (Jenkins & Duncan, 2017; Phillips et al., 2017). Prior studies on vocabulary growth, in particular, have found positive effects of coupling reading and playing (Han, Moore, Vukelich, & Buell, 2010; Hassinger-Das et al., 2016; Roskos & Burnstein, 2011; Weisberg, Kittredge, Hirsh-Pasek, Golinkoff, & Klahr, 2015). Indeed, Lillard et al. (2013) suggested that one of the few areas where play had a demonstrable effect was in language learning. What prior studies have not explored is the comparative and unique effects of different types of postreading play activities on preschoolers' word learning. Our research addresses this gap.

Specifically, our three different approaches in Study 1 were free play and two approaches involving adult participation and support: guided play and directed play. Free play, which grants children freedom for exploration, allows them to engage in discovery learning and practice skills with no constraints from adult involvement (Fisher, Hirsh-Pasek, Golinkoff, Singer, & Berk, 2010; Johnson, Christie, & Yawkey, 1999; Singer & Singer, 1990). Pyle and Danniels (2017) found free play to be the most common form of play in early childhood classrooms. This approach to play capitalizes on children's interests and provides a meaningful context in which to encounter vocabulary, thus tapping into at least two of the six principles of vocabulary acquisition (Harris et al., 2011; Hassinger-Das et al., 2017; Konishi et al., 2014).

Although we included a Free Play condition in Study 1 to see whether it could facilitate vocabulary learning, the literature suggests that free play is not an optimal pedagogy when there is a specific learning goal (Chien et al., 2010; Fisher, Hirsh-Pasek, Newcombe, & Golinkoff, 2013; Klahr & Nigam, 2004). Instead, adults' subtle assistance through *guided play* has been associated with better learning outcomes. Guided play falls under the umbrella of playful learning (Hirsh-Pasek & Golinkoff, 2011; Hirsh-Pasek et al., 2009; Weisberg, Hirsh-Pasek, Golinkoff, Kittredge, & Klahr, 2016), preserving the playfulness and child-directedness of free play and the goal-directedness of formal instruction (Ash & Wells, 2006; Han et al., 2010; Rogoff, 2003; Roskos, Tabors, & Lenhart, 2004; Toub, Rajan, Golinkoff, & Hirsh-Pasek, 2016; Weisberg, Hirsh-Pasek, & Golinkoff, 2013; Weisberg et al., 2016). Guided play builds on theories emphasizing the significance of social interactional contexts for children's learning. According to these frameworks, a significant portion of children's learning happens in shared meaningful contexts, through apprenticeships that occur against the backdrop of guided interaction (Rogoff, 2003; Vygotsky, 1978). In this spirit, adults using guided play for vocabulary support find opportunities to infuse children's unfolding play with references to target words, reminders of their meaning, and questions that link the meaning to prior knowledge and that invite them to apply and expand upon their knowledge. Because it exemplifies all six principles of vocabulary acquisition (Harris et al., 2011; Hassinger-Das et al., 2017; Konishi et al., 2014), we anticipated that guided play would be a successful strategy.

Our third approach to play was *directed play*. Here, the adult leads the play as an external director, rather than serving as a supportive participant. For example, directed play might involve an adult leading children in a re-enactment of a storybook using figurines. This approach resembles direct instruction methods in which the adult explains the exact meaning of concepts and controls the flow of behavior (Kirschner, Sweller, & Clark, 2006). Directed play also embodies all the principles of vocabulary acquisition except possibly the interactive and responsive principles of language learning, since there is not much room for consideration of children's contributions (Harris et al., 2011; Hassinger-Das et al., 2017; Konishi et al., 2014). Directed play might be an effec-

tive combination of an element of playfulness with a more didactic approach. However, this type of play is not playful learning, because children do not lead.

Here, we empirically compare the benefits of these three different play approaches for post-reading vocabulary learning. By first testing playful vocabulary review through school visits by members of the research team (Study 1) and then shifting to delivery by classroom teachers (Study 2), we examine the feasibility and effectiveness of these methods in practical settings. Additionally, our Study 2 comparison of a playful approach versus a more didactic approach further unpacks the unique benefits of a playful pedagogy. By starting with the established activity of book-reading, these studies probe the possible *added* value of play as a word learning tool.

2. Study 1: test of three play-based approaches

In Study 1, children experienced an enriched shared book-reading activity followed by either free play, guided play, or directed play. We investigated the following research questions: (1) Do children in this intervention show improvements in their knowledge of the targeted vocabulary over time? (2) Which approach to play is most effective for supporting such vocabulary development? Additionally, since maintenance of gains is part of successful learning (Neuman, Newman, & Dwyer, 2011), we took advantage of an opportunity at one of our sites to explore children's retention of word knowledge after a delay of two weeks.

We hypothesized that, across conditions, children would have greater receptive and expressive knowledge of target words at post-test than at pre-test. Gains were also expected to be greater for target words than for words that were present in the story but not explicitly taught (exposure words) and for words that were not included at all (control words). We also hypothesized that the two conditions involving adult support for targeted learning during play (Guided and Directed Play) would promote significantly greater receptive and expressive vocabulary growth than would the Free Play condition in which the adult only provides toys. Further, Guided Play was expected to be the strongest condition because it incorporated more of the language learning principles. Finally, we hypothesized that gains would be sustained after a delay of two weeks. While our investigation of retention was more exploratory in nature, we suspected that the Guided Play condition would show the greatest knowledge after the additional delay, since the learning was hypothesized to be stronger in that condition.

Echoing others' efforts to better understand how intervention programs can impact subgroups of children differently (Greenberg & Abenavoli, 2017; Miller, Farkas, Vandell, & Duncan, 2014), we collected additional background information to address possible moderators of the effects of different approaches to play. Specifically, based on prior research on varied language learning trajectories, we gathered data on children's home languages and English Language Learner status (Hindman & Wasik, 2013; Hoff, 2013; Miller et al., 2014), and maternal education (Dollaghan et al., 1999; Hindman & Wasik, 2013; Qi, Kaiser, Milan, & Hancock, 2006; Richels, Johnson, Walden, & Conture, 2013).

3. Method

3.1. Participants

After obtaining approval from our Institutional Review Boards (IRBs), participants were recruited from 10 Head Start preschool classrooms in Eastern Pennsylvania and from 18 Pre-K classrooms in Central Tennessee. The Head Start program serves only low-

income families, and Tennessee's program prioritizes enrollment of economically disadvantaged children.

Several weeks before the start of the study, classroom teachers distributed consent forms (in English and Spanish) to children's families. From the two sites, consent was provided for 258 children (165 in Tennessee). Nine children withdrew from the study and were excluded from analyses. The final sample of 249 included approximately nine children per classroom (46% male; M_{age} = 59.19 months, SD_{age} = 4.75 months, range$_{age}$ = 39–66 months).

Within each classroom, consented children were first randomly assigned to condition. Then, and 3–4 children from the same condition were assigned to a small group to participate in the activities together. Most classrooms only had one group per condition, and teachers were kept blind to the groups' conditions. Adjustments to these initial arrangements were made in a small minority of cases, based on consideration of teachers' input on behavioral incompatibility, an aim to have mixed gender groups, and attendance. These adjustments were made blind to children's pre-test abilities. Distribution across condition was almost exactly equal, with 84 children in Free Play, 83 children in Guided Play, and 82 children in Directed Play. As a preliminary way of exploring retention of gains after an additional delay, we conducted follow-up testing at the Tennessee site (n = 153; 7 children were absent). This exploration of retention was not an original focus, as were our other research questions; it was a bonus set of data that was possible at only the Tennessee site, and we elected to incorporate it as an additional, secondary exploration.

All but three children previously participated in a prior study in this XXXX project, which compared shared book-reading approaches, followed by a related free play session. In that study, children's vocabulary words and books were different from the ones they encountered in this study. No participants had referrals or Individualized Education Programs (IEPs) relating to cognition or language, and all children had sufficient vision, hearing, and English proficiency. Teacher report indicated that 14.9% of the sample was designated as English Language Learners (ELL; n = 37), and parent report indicated that Spanish was the home language for the majority of those children (see Table A1 for additional information about children's languages). Race and ethnicity information was provided for all but 5 of the final participants (55.3% African-American; 23.4% Hispanic/Latino, 0.8% Asian, 13.5% Caucasian, and 7.0% Other or Multiracial). Parents reported highest level of maternal education for 196 children: 17.3% some high school, 41.3% high school diploma or GED, 18.4% trade school, 13.3% associate's degree, 6.1% bachelor's degree, and 3.6% graduate degree (see Table A2 for additional family background data). Comparison across the two sites indicated statistically significant differences in children's age, ELL status, home language, maternal education, and race/ethnicity. Moderation of condition effects by site was therefore examined in our analyses.

3.2. Materials, intervention, and measures

3.2.1. Book and vocabulary word selection

Two theme-based pairs of commercially available books were selected after confirming that they were not part of the participating classroom's library collections: dragon, with *The Knight and the Dragon* (dePaola, 1998) and *Dragon for Breakfast* (McMullen & McMullen, 1990), and farm, with *Farmer Duck* (Waddell & Oxenbury, 1991) and *Pumpkin Soup* (Cooper, 2005).

The design involved three types of vocabulary words (see Table A3): (1) *target words* (10 per book; 20 per theme), which were explicitly taught with specific strategies during the intervention activities; (2) *exposure words* (3–5 per book; 8–9 per theme), which were mentioned in the text of the book and sometimes during play

but never explicitly taught; and (3) *control words* (8 per theme), which were not included in any intervention materials or activities. We tested exposure and control words to isolate learning that was the result of our teaching efforts and not due to general vocabulary growth. The original books contained some words that were used for target and exposure words, and we also inserted additional words into the texts, as needed. In adapting the texts, we ensured the desired frequency of 1–2 usages per target vocabulary word and similarities across books in total number of words, number of words per page, total number of pages, and pictorial representations of target words. The adaptations did not significantly alter the storylines from the original books; they typically described additional details.

In selecting target, exposure, and control words, we sought adult-conversational words of high utility, as in the "Tier 2" designation by Beck, McKeown, and Kucan (2013). Word selection was also driven by feasibility of child-friendly definitions, semantic and phonological distinctness, difficulty, frequency of use (Biemiller, 2010; Chall & Dale, 1995), and the likelihood of words' unfamiliarity, based on prior data from this project. Each set of words contained concrete and abstract nouns, verbs, adjectives, and spatial prepositions, since children's word-learning might vary based on form class and concreteness (Beck & McKeown, 2007; Hadley et al., 2016).

3.2.2. Intervention procedure

Data were collected March–May 2012. Participating classrooms were randomly assigned to the dragon or farm theme during a prior study, and assignments were swapped for this phase so that participants encountered a new theme, with new books and new words. Groups of 3–4 children assigned to the same condition engaged in our activities with an Intervention Specialist (IS). These ISs were recruited through advertisements in the community (e.g., on a university job listings site and at community colleges) which targeted people who were interested in children's language development and who had previously worked with young children. All nine ISs were females with experience in early childhood settings and had or were completing related degrees. Some had previously worked as preschool teachers, and others had worked at local libraries or book stores reading books to children. ISs completed 2–3 training sessions (3–4 h each) in which they reviewed the scripts and guidelines with the research team and participated in role-playing exercises to practice leading the various book-reading and play sessions. ISs conducted book-reading sessions and all three types of play sessions in their assigned classroom(s). Most ISs had 8 play sessions of each type (Free Play, Guided Play, and Directed Play; 4 sessions per group for each of 2 books in the theme).

During each of two weeks, children participated in four consecutive days of back-to-back book-reading and play sessions. Activities were based on the first book in the theme in week 1 and the second book in the same theme in week 2, with half of the groups that were assigned to a given theme starting with one book and half starting with the other. ISs' records showed that children attended the small group activities an average of 7.14 of the 8 intervention days (SD = 1.30, range = 3–8 days), with no condition differences and no opportunities to make up absences. There were two main measurement periods: pre-testing occurred within one week prior to the first day of intervention activities, and post-testing occurred within one week after the last day of activities. A follow-up round of testing occurred with a subsample of children two weeks after post-testing. Testing typically occurred by classroom, with 2 days per classroom per measurement period. Therefore, within each classroom, children from all conditions were tested within the same 2-day period, and any differences between classrooms in the amount of time that passed between activities and testing periods were not confounded with condition comparison. On average

23.78 days ($SD = 6.95$) elapsed between pre- and post-testing. Time elapsed was not significantly related to post-test scores.

3.2.2.1. Book-reading. For each book, we created a script to guide ISs through each 10-min reading, and ISs were instructed to minimize deviations from the scripts. The script outlined evidence-based strategies (Biemiller & Boote, 2006; Goldin-Meadow & Alibali, 2013; Sénéchal, 1997; Wasik et al., 2006) for enriching the book-reading by elaborating on the book's text and illustrations. Each of a book's 10 target words was included in the book's text. Direct instruction of word meanings was supplied across all readings; children were encouraged to participate more in later readings. Initially children were given a child-friendly definition, commentary that included additional conceptual information about the word's meaning, encouraged to use a gesture, and were directed to look at relevant parts of the pictures. For example, when the target word **fierce** appeared in the book, the IS used her script to explain, "The dragon wants to scare the knight so he is practicing making a **fierce** face [point to picture] [growly scary voice]. He shows his teeth like this [make a face]. You try to make a **fierce** face that might scare someone. You guys look scary!" Scripts for later sessions encouraged children to produce the words by first giving a brief review of key content from the page and then pausing at the point the word was to be used. The brief review supplied additional information about the words, and prompts for gestures were included when relevant: "The dragon also wanted a book on how to fight. What did he do? Yes, he looked around for a book on how to fight knights; he **rummaged** around in his box for a book. Let's pretend we're **rummaging** around to look for a toy." After the reading, the IS reviewed the story and words using illustrations from the book and separate picture cards to briefly review target words. These reviews also gradually increased demands on children to jointly construct the story and produce the words themselves: "The dragon also tried to scare the knight by showing his sharp claws. What do we call an animal's claws? Yes, **talons**. [show picture card] Let's all say that now! **Talons**. [children repeat]] What else can animals do with them? [catch food]." Comprehension-related discussions were included in the scripts for each book reading, with progressively more challenging questions as the story became more familiar.

3.2.2.2. Play. There was a collection of toys related to each book, with a setting piece (i.e., castle, farm house), figurines, and props relevant to the story. One complete set of toys was available to the groups of children for a 10-min play period following each book-reading. In the Free Play condition, children played with the toys in whatever manner they chose, with the IS supervising and only intervening regarding safety or conflicts. By design, this condition did not involve active adult support of vocabulary. In the other two conditions, the IS actively participated in the play and added reviews of target vocabulary words. To minimize task demands, only half of each book's 10 target words were designated as focus words on a given day, so each word was reviewed during two play sessions.

In the Directed Play condition, the IS followed scripted language to direct children in a reenactment of the book's story. Each script included three instances of each of the day's target words: a brief definition and two uses as part of the IS's narrative and re-enactment directions (e.g., "...everyone in the castle [was] so surprised to see a dragon *emerging* from the egg," "*Emerging* means to come out of something," and "Can you make the dragon *emerge* from, or come out of, the egg?"). If necessary, ISs could adapt the script based on their interactions with children, but they were instructed to follow the script as closely as possible to incorporate the words properly. Children were expected to listen to the IS's narrative and play along.

In the Guided Play condition, children chose what to do during their play. The IS did not read from a script or require children to follow re-enactment instructions. Instead, she was told to follow children's lead, join their play (e.g., manipulating the toys and talking as a character), and incorporate target vocabulary as naturally as possible. To keep frequency of exposure controlled across adult-supported play conditions, the IS was instructed to incorporate each of the day's five assigned words three times, just like in Directed Play. In this condition, however, the three instances included a definition, a closed-ended question, and an open-ended question. Closed-ended questions had easy answers, often based on the question itself, the word's definition, or the story. Open-ended questions challenged children to think more deeply about the word, and there was no single correct answer. The aim was to find natural opportunities to incorporate the words into the more child-directed play. Instead of a verbatim script, a guidance card outlined key elements of the words' definitions (e.g., "coming out of something" for *emerging*) and sample questions (e.g., "Does *emerging* mean going into or coming out of?" for a closed-ended question or "Can you think of some other things [character] can *emerge* from?" for an open-ended question). The IS kept the assigned words in mind and looked for opportunities to build on what children were already doing and to segue into reviewing a related vocabulary word.

3.3. Measures

Children were tested one-on-one (20–25 min) with both a receptive and an expressive test, in a counterbalanced order. All words associated with a given theme were tested on at least one, if not both, of these vocabulary tests.

3.3.1. Receptive test

We devised a receptive test similar to the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007), but with fewer foils and incorporating the specific words in our study. The experimenter displayed a page with three images in a row and asked the participant to point to the image showing a representation of a particular word (e.g., "Where's *throne*?"). Each page included the correct referent, a conceptual foil, and a thematic foil (e.g., throne, folding chair, and crown, respectively). All images differed from the books' illustrations, the picture cards used during book reading, and play session toys, making this a test of extension and not just recognition of familiar images. Positions of the three types of images within the row varied. Children were tested on their theme's words: 20 target words (from the two books), 8 control words, and as many of the 8–9 exposure words as could be depicted (6 for farm, 8 for dragon). Children's scores reflected the percentage of test items for which they chose the correct image. Internal consistency for the receptive measure was calculated from children's post-test data, where we expected to see above-chance responses. With Cronbach's $\alpha = .72$ for farm and .61 for dragon, results were acceptable but on the low side, perhaps partially due to the 33% chance of being right on a given item simply by guessing and to the narrow range of scores within our sample.

3.3.2. Expressive test

We used the New Word Definition Test-Modified (NWDT-M; Hadley et al., 2016), modeled after Blewitt, Rump, Shealy, and Cook (2009). Children were asked to explain the meaning of vocabulary words using both verbal response and gestures. The experimenter asked children one question (e.g., "What are *talons*?") and then prompted them further, after their initial response (e.g., "Can you tell me or show me anything more about *talons*?"). To keep typical task duration under 15 min, the expressive test contained a subset of words from our study—21 total for dragon (13 target, 4 exposure, 4 control) and 18 total for farm (8 target, 6 exposure, 4 control).

There were three exposure words—tip, plan, accidentally—that were not featured on the receptive test, because they were not easy to depict visually. These words were included in the expressive test, along with a sampling of target, exposure, and control words from other form classes (no less than 67% of any form class of word). Video- or audio-recordings were used to facilitate later scoring.

3.3.2.1. Scoring. A child's expressive task score was the average information units the child provided per word. Information units were indications of knowledge about the word's meaning, including the use of synonyms, antonyms, part-whole or hierarchical relations, function, perceptual features, gestures, and meaningful context (each counting for 1 point) or basic context (counting for 0.5 point because it involved an association without semantic information, such as only saying "Santa Claus" for the word chimney). For example, for throne, "a chair that queens sit on" has 3 information units: "chair" is a synonym, and "that queens sit on" provides meaningful context and describes function. The range of possible scores begins at 0 (no information units provided for any item) but has no predetermined maximum, since participants are not limited in the amount of information that they can provide in their responses to the open-ended prompts. Scoring for the expressive task is described in more detail in Hadley et al. (2016).

Four masters' students in education served as expressive task coders, and they demonstrated at least 90% agreement with the doctoral student Gold Standard coder during training. Then, every fifth set of child responses consecutively coded by an individual coder was also compared against the Gold Standard coder's data to prevent coder drift. In the case of a discrepancy, the Gold Standard coder's code was used, and the other coder re-calibrated through additional coding practice and feedback before continuing. The average percent agreement for the 20% of children scored for reliability was 93.2% (Cohen's kappa = 0.82), indicating substantial agreement.

3.3.3. Parent questionnaire

A brief questionnaire (available in English and Spanish) about family background was sent home with participants. Families of 242 participants completed all or most of the questionnaire items. For the purposes of addressing potential moderator effects for our second research question, we used parent questionnaire data to create composite scores based on principal component analysis (PCA) of relevant questionnaire items. For example, children's home literacy exposure was a composite of items asking if the parent has a library card (yes/no), how many books the child has (<5, 6–10, 11–20, or >20), and how many times per week someone reads to the child (0–1, 2–3, 4–5, 5–6, or every day). Home language was similarly a composite of responses regarding the proportions of time the child hears and speaks English and other languages. Another family background item was maternal education, documented as either some high school, high school diploma or GED, trade school, associate's degree, bachelor's degree, or graduate degree. With the median falling at a high school diploma (or GED), a dichotomous split was between those with a maximum of a high school diploma (n = 113) and those with further education (n = 83).

3.4. Fidelity of implementation

Researchers (e.g., graduate students and postdoctoral fellows involved in the study) coded a subset of the book-reading and play sessions by counting the percentage of assigned words that the IS reviewed within the session; they then calculated an average score for each IS. Coders attended at least one of each IS's book-reading sessions and provided feedback to the IS upon completion of the session or soon afterward. Observations indicated that each IS reviewed at least 90% of the 10 target words during the book-reading sessions, which were scripted. Members of the research team at both sites observed at least one Guided and one Directed Play session for each IS (ISs were uninvolved in Free Play). Researchers listened for ISs' reviews of the 5 assigned words for that day. All ISs scored at least 90% on Guided Play and on Directed Play, and they were given feedback regarding errors. To further confirm fidelity during Guided and Directed Play sessions, researchers coded recordings of 4 sessions of each type from each IS. For Guided Play, each IS at both sites scored an average of at least 90% across 4 coded sessions, with half of the ISs scoring 100% on all 4 sessions. The lowest single-session score was 80%. For Directed Play, the Pennsylvania site had technical problems that left all but one IS's recordings unusable, but that one IS scored 100% on all 4 coded sessions. At the Tennessee site, all but one IS scored an average of at least 90%, and the remaining IS had an average of 80% across 4 sessions, due to one initial outlier session. That IS was given feedback, and she immediately improved, with a 90% average across all 8 Directed Play sessions.

3.5. Data analyses

Using multilevel modeling (MLM) to account for interdependency (children nested within intervention play groups, nested within classrooms), we examined the effectiveness of each play approach for improving children's vocabulary knowledge. Fewer than 5% of children were missing data on one of the two study outcomes, and analyses were conducted on all available cases for a given outcome. Unless otherwise noted below, all posthoc pairwise comparisons were conducted using Fisher's Least Significant Difference (LSD) test, and all effect sizes are presented as Cohen's d (Cohen, 1988).

4. Results

Table 1 shows mean raw scores and standard deviations for all measures by time for the full sample of children and the subsample that was followed up after an additional delay.

4.1. *Children's* vocabulary gains over time

Our first research question asked if children gained vocabulary knowledge through this intervention. In separate models for the two vocabulary tests, $time_t$ was nested within $children_i$ and differences between time points were explored (γ_{10}) (see Supplementary Materials, Eq. (1)). Collapsing across condition, initial analyses showed that children's increase in scores from pre-test to post-test were statistically significant ($ps < .01$) for both receptive and expressive knowledge of target ($B_{receptive}$ = 0.21, SE = 0.01, Cohen's d = 1.34; $B_{expressive}$ = 0.42, SE = 0.03, d = 1.34), exposure ($B_{receptive}$ = 0.06, SE = 0.01, d = 0.56; $B_{expressive}$ = 0.10, SE = 0.02, d = 0.26), and control words ($B_{receptive}$ = 0.05, SE = 0.01, d = 0.28; $B_{expressive}$ = 0.06, SE = 0.02, d = 0.24).

Tests of baseline equivalence among target, exposure, and control words indicated that children showed the least knowledge of target words at pretest and the most knowledge of control words ($p < .05$). To account for potential differences at baseline, our model controlled for vocabulary knowledge at baseline. Using that model, we next examined if children learned target words better than exposure and control words. As seen in the top panel of Table 2, pre-to post-test receptive and expressive gains for target words were significantly larger than gains for exposure words ($d_{receptive}$ = −0.32 and $d_{expressive}$ = −0.47) and for control words ($d_{receptive}$ = −0.58 and $d_{expressive}$ = −0.89). Posthoc pairwise comparisons also indicated that gains for exposure words were significantly larger than gains

Table 1
Study 1: receptive and expressive vocabulary unadjusted means (standard deviations) by assessment session for full sample and subsample with follow-up data.

Measures	Full sample ($N = 249$)		Subsample with Follow-up data ($n = 153$)		
	Pre-test	Post-test	Pre-test	Post-test	Follow-up
Receptive and expressive vocabulary by level of instruction (collapsed across conditions) (N = 249)					
Target words					
Receptive vocabulary	0.37 (0.13)	0.58 (0.17)	0.37 (0.13)	0.59 (0.17)	0.61 (0.18)
Expressive vocabulary	0.14 (0.21)	0.56 (0.48)	0.15 (0.24)	0.60 (0.50)	0.60 (0.49)
Exposure words					
Receptive vocabulary	0.50 (0.20)	0.56 (0.22)	0.51 (0.18)	0.57 (0.23)	0.61 (0.24)
Expressive vocabulary	0.38 (0.42)	0.48 (0.26)	0.40 (0.42)	0.48 (0.44)	0.52 (0.50)
Control words					
Receptive vocabulary	0.45 (0.17)	0.51 (0.18)	0.45 (0.17)	0.51 (0.19)	0.50 (0.19)
Expressive Vocabulary	0.19 (0.23)	0.25 (0.26)	0.21 (0.24)	0.25 (0.25)	0.29 (0.31)
Target receptive and expressive vocabulary by experimental condition					
Guided play					
Receptive vocabulary	0.37 (0.12)	0.60 (0.17)	0.38 (0.11)	0.62 (0.17)	0.62 (0.17)
Expressive vocabulary	0.14 (0.21)	0.59 (0.48)	0.16 (0.24)	0.67 (0.50)	0.64 (0.52)
Directed play					
Receptive vocabulary	0.37 (0.13)	0.60 (0.16)	0.38 (0.14)	0.63 (0.15)	0.66 (0.16)
Expressive vocabulary	0.16 (0.24)	0.64 (0.51)	0.17 (0.27)	0.70 (0.53)	0.66 (0.48)
Free play					
Receptive vocabulary	0.37 (0.13)	0.54 (0.18)	0.37 (0.13)	0.54 (0.18)	0.55 (0.21)
Expressive vocabulary	0.11 (0.17)	0.42 (0.41)	0.12 (0.19)	0.44 (0.44)	0.50 (0.47)

Note. The full sample was distributed across Guided Play ($n = 83$), Directed Play ($n = 82$), and Free Play ($n = 84$), and the subsamples with follow-up data had $n = 52$, $n = 51$, and $n = 50$, respectively. Receptive task values indicate the proportion of items that were answered correctly, with chance level equal to 0.33 (i.e., select one of three images). Expressive task values indicate the average number of information units children provided per expressive item, and these scores ranged from 0.00 to 2.63. For comparison, the range for Hadley et al. (2016) was 0–2.02.

Table 2
Study 1: parameter estimates (standard errors) for effects of level of instruction (top panel) and condition (middle panel) at post-test and effect size for condition effects (bottom panel).

Parameters	Receptive vocabulary	Expressive vocabulary
Level of instruction effects		
Level 1, level of instruction		
Intercept, γ_{000}	0.392 (0.023)	−0.500 (0.202)
Pre-test performance, γ_{100}	0.335 (0.039)**	0.617 (0.042)**
Target versus exposure, γ_{200}	−0.062 (0.015)**	−0.220 (0.029)**
Target versus control, γ_{300}	−0.102 (0.015)**	−0.342 (0.027)**
Condition effects (full sample)		
Level 1, child		
Intercept, γ_{000}	0.180 (0.134)	−0.748 (0.341)
Pre-test performance, γ_{100}	0.473 (0.078)**	1.229 (0.124)**
Age at pre-test, γ_{200}	0.001 (0.002)	0.018 (0.005)**
Attendance, γ_{300}	0.026 (0.008)**	0.031 (0.019)
Gender (males = 0), γ_{400}	0.007 (0.020)	0.024 (0.050)
Level 2, play group		
Guided play versus directed play, γ_{010}	−0.010 (0.023)	0.035 (0.060)
Guided play versus free play, γ_{020}	−0.068 (0.023)**	−0.116 (0.059)*
Level 3, classroom		
Theme (0 = farm), γ_{001}	−0.028 (0.021)	−0.212 (0.053)**
Cohen's d effect sizes		
Guided play versus directed play, γ_{010}	−0.061	0.069
Guided play versus free play, γ_{020}	−0.388	−0.260
Directed play versus free play	−0.341	−0.324

Note. $N = 249$. Standard errors adjusted for interdependency among observations. For test of Level of Instruction effects, target words is the reference group for the comparison (negative estimates indicate that target words had larger covariate adjusted post-test scores). For test of condition effects, guided play condition is the reference group for the comparison (negative estimates indicate that children in the guided play condition had larger covariate adjusted post-test scores compared to children in the comparison condition). Positive estimates for gender indicate that females made greater gains compared to males. Positive estimates for Theme indicate that children in the Farm theme had larger post-test gains compared to children in the Dragon theme. Contrast of Directed Play (comparison group) versus Free Play was estimated with post-hoc pairwise comparison.

  * $p < .05$.
  ** $p < .01$.

for control words on both receptive ($p = .005$, $d = −0.20$) and expressive measures ($p < .001$, $d = −0.33$).

4.2. Comparison of play conditions

Our second research question focused on comparative benefits of the three different play conditions for children's gains in vocabulary knowledge. Tests of baseline equivalence for receptive and expressive pre-test scores, gender, age, and ELL status were conducted in separate multilevel models, with each variable being predicted by condition assignment. Results indicated that pre-test differences among the conditions were non-significant ($ps > .10$). Based on the means provided in Table 1, the magnitude of the difference was zero across conditions for the receptive measure. For the expressive measure, children in the Free Play condition tended to have lower scores at pretest ($d = 0.16$ for contrast with Guided Play and 0.24 for Directed Play), while the difference between Guided and Directed Play conditions was small and non-significant ($d = 0.09$). Baseline vocabulary (γ_{100}), age (γ_{200}), gender (γ_{300}), attendance (γ_{400}), and theme (γ_{001}) were included as covariates in models examining condition effects (γ_{010} and γ_{020}) on vocabulary gains, which incorporated children$_i$ nested within playgroup$_j$ and classroom$_k$ (see Supplementary Materials, Eq. (2)).

In this discussion of condition effects, we present results for target words only, as those are the only words that were systematically treated differently by condition. Results for both the receptive and expressive measures (see Fig. 1) indicated that children in Guided Play made significantly larger gains than did children in Free Play, $d = −0.39$ and $−0.26$, respectively (see Table 2, middle and bottom panels). Similarly, posthoc comparisons indicated that children in Directed Play made significantly larger receptive and expressive gains than children in Free Play ($ps < .015$), $d = −0.34$ and 0.32, respectively. There were no differences between children in the Guided and Directed Play conditions for receptive and expressive gains, $d = −0.06$ and 0.07, respectively.

To explore whether the comparative benefits of the different play conditions would vary based on child characteristics, we conducted a series of analyses to examine if children's home literacy
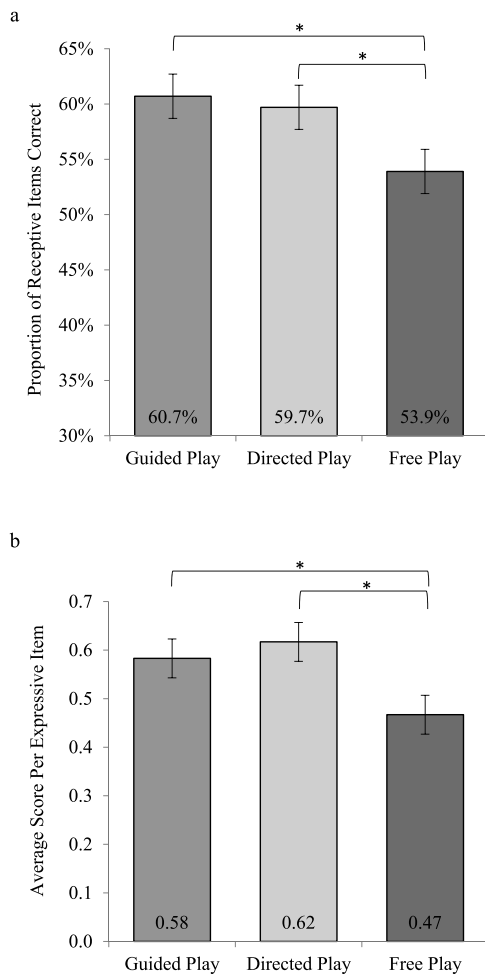
a

b

Fig. 1. Study 1 vocabulary gains by condition. Proportion of items correct on the receptive measure of vocabulary (a) and the average score per expressive item (b) at the post-test controlling for pretest performance, attendance, age, gender, and theme. Standard errors represented in the figure by the error bars attached to each column. *Condition contrast significant at $p < .05$.

exposure (composite of responses regarding library card, number of books, and reads per week), maternal education, home language (composite of what children hear and speak at home), or ELL status moderated condition effects for receptive or expressive post-test scores (appropriate interaction terms were added to Eq. (2)). Analyses involving home literacy exposure, home language, and maternal education were conducted only on 80% of the sample ($n = 195$ or 196 depending on analysis), as some children were missing those parent-report data. Based on What Works Clearinghouse attrition standards (2014), examination of the reduction in sample sizes across condition indicated that missing data rates were approximately equivalent across condition (less than 0.08 SD). We also tested for moderating effects of children's pre-test scores (to address potential Matthew effects), attendance during our activities (to address dosage), and site (to address sample, implementation, or other differences across sites).

The only statistically significant results were in the models testing the dichotomous maternal education variable as a moderator of receptive and expressive vocabulary gains (see Table A4). For receptive vocabulary, maternal education moderated condition differences between Guided Play and Free Play ($B = 0.15$, $SE = 0.06$, $p = .01$) and Directed Play and Free Play ($B = 0.12$, $SE = 0.06$, $p = .03$) but not between Guided and Directed Play. Children whose mothers had higher education showed greater gains in Guided Play compared to Free Play ($p < .01$, $d = 0.73$) and in Directed Play compared

to Free Play ($p < .01$, $d = 0.76$). For children whose mothers had lower education, there were no condition effects.

For expressive vocabulary, maternal education moderated differences between Guided and Free Play ($B = 0.34$, $SE = 0.13$, $p = .01$) but neither of the other comparisons. Children of mothers with higher education showed greater gains in Guided Play than in Free Play ($p < .01$, $d = 0.62$). There were no condition effects for children of mothers with lower levels of education.

4.3. Follow-up testing results

We addressed the question of retention through data from the subsample of children who participated in follow-up testing two weeks after the post-testing. Expanding upon the model to test condition differences (see Supplementary Materials, Eq. (2)), tests of follow-up effects included an additional nesting level to account for observations$_t$ (γ_{100} and γ_{200}) nested within children$_i$ nested within play group$_j$. Of primary interest were the cross-level interactions between condition and observation (γ_{101}, γ_{102}, γ_{201}, and γ_{202}) (see Supplementary Materials, Eq. (3)). For these children, gains and condition effects between pre- and immediate post-test were similar to those for the full sample reported above (see Table 3). The results from further analyses indicated that there was no statistically significant change in children's vocabulary knowledge from post-test to follow-up for either the receptive or the expressive measure. This pattern in the data means that initial gains made over the course of the intervention were preserved for at least another couple of weeks. This also means that effect sizes for condition differences at the follow-up timepoint were comparable to those reported above for the full sample's pre- to post-test gains. More specifically, for both the receptive and expressive measures, children in Guided Play had significantly higher follow-up scores than did children in Free Play, $d = -0.28$ and -0.28, respectively. Similarly, children in Directed Play had significantly larger follow-up scores than did children in Free Play for receptive and expressive gains, $d = -0.33$ and -0.24, respectively. There were also no differences between children in the Guided and Directed Play conditions for receptive and expressive gains, $d = 0.04$ and 0.05, respectively. Moderators were not examined for the follow-up data due to the reduced sample size and the similarity between post-test and follow-up.

5. Discussion of study 1

The overarching purpose of this project was to teach low-income preschoolers new words through a combination of book-reading and play. Our first main research question for Study 1 was whether preschoolers taught words through book reading and play could learn those target words better than they would learn exposure and control words. Our findings, controlling for many variables, indicate that the vocabulary instruction we offered was effective: The intervention improved children's abilities to recognize novel depictions of the target words *and* to explain words' meanings. This is a valuable finding, as transfer and depth of knowledge are often not well supported in pre-K curricula (Neuman & Dwyer, 2009).

Our second main research question was whether variations in adult support for vocabulary learning during play would make a difference in word learning. The results support our hypothesis that children in the adult-supported play conditions (Guided Play and Directed Play) would outperform their peers in the Free Play condition on both the receptive and expressive measures. One likely reason for poorer learning in Free Play is the lack of built-in vocabulary review for the target words. It was important to explore this type of play, however, because it most closely approximated the

Table 3
Study 1: parameter estimates (standard errors) for models estimating condition differences for receptive and expressive vocabulary gains (post-test & guided play = reference groups).

Parameters	Receptive vocabulary	Expressive vocabulary
Fixed effects		
Level 1, observation		
Intercept, γ_{000}	0.188 (0.204)	−0.727 (0.574)
Post-test versus pre-test, γ_{100}	−0.249 (0.024)**	−0.515 (0.055)**
Post-test versus follow-up, γ_{200}	0.007 (0.024)	−0.035 (0.055)
Level 1, child		
Age at pre-test, γ_{010}	0.004 (0.003)	0.020 (0.009)**
Attendance, γ_{020}	0.022 (0.011)	0.024 (0.030)
Gender, γ_{030}	0.016 (0.022)	0.050 (0.061)
Level 3, play group		
Guided versus directed, γ_{001}	−0.001 (0.033)	0.036 (0.086)
Guided versus free, γ_{002}	−0.080 (0.033)*	−0.242 (0.086)**
Theme (0 = farm), γ_{003}	−0.007 (0.022)	−0.003 (0.060)
Cross-level interactions		
Guided v. directed * post v. pre, γ_{101}	0.008 (0.033)	−0.012 (0.077)
Guided v. directed * post v. follow-up, γ_{201}	0.020 (0.033)	0.006 (0.077)
Guided v. free * post v. pre, γ_{102}	0.073 (0.033)*	0.197 (0.076)**
Guided v. free * post v. follow-up, γ_{202}	−0.001 (0.033)	0.090 (0.077)

Note. N = 153. Standard errors adjusted for interdependency among observations. Observations (Level 1) were repeated within children (Level 2), who were nested within Play Groups (Level 3). The guided play condition is the reference group for the comparison of condition difference; as such, negative estimates indicate that children in the guided play condition had larger post-test scores compared to children in the comparison condition. Post-test is the reference group for the comparison of observation effects; as such, negative estimates indicate that scores at post-test were larger than the comparison observation. Positive estimates for Theme indicate that children in the Farm theme had larger post-test scores compared to children in the Dragon theme.

* $p < .05$.
** $p < .01$.

style of play used in many classrooms. In terms of play with adult support, we found no statistically significant differences between the Directed and Guided Play groups on vocabulary gains. Although we had expected some additional benefit in Guided Play, the less child-directed play within the Directed Play condition was similarly effective for vocabulary support.

An intriguing finding is the fact that these two play conditions' effects were moderated by the variable of maternal education. That is, children's vocabulary gains in both Guided and Directed Play were beneficial only for the subset of children whose mothers had greater than a high school diploma. Children in our study whose mothers had less education might not have had as much experience with learning through adult-supported contexts and therefore did not benefit as much from adult support. Consistent with this interpretation, findings from a recent article (Hirsh-Pasek, Adamson et al., 2015) show variations even within a low-income sample in the quality of mother-infant interactions during book-reading and play activities. In that sample, a measure of fluency and connectedness between mother and infant was significantly and positively correlated with maternal education. Given that maternal education moderated children's learning here, for Study 2 we sought to develop a new approach that might better support more children by combining elements from the two adult-supported conditions.

As for the issue of retention, children at the site where follow-up testing was conducted retained their newfound knowledge after an additional two weeks, and the same overall pattern emerged, with a relative benefit of Guided and Directed Play conditions compared to the Free Play condition. This result illustrates the strength of our approach. If children do not retain vocabulary gains, they cannot benefit from either the knowledge of those specific words or the snowball effect in which knowledge of some words can promote understanding of other new words and concepts (Neuman et al., 2011). These data are from only one of the two sites, and it was beyond the scope of this study to conduct extended follow-up testing (i.e., after an additional 3–6 months) or to explore any longer-term effects on general vocabulary growth; however, these initial follow-up results suggest that using adult-supported play for vocabulary review is a promising avenue for further pursuit.

6. Study 2: play versus non-play vocabulary review activities

We have argued that it was the play-based adult support for word-learning that led children in the Guided and Directed Play conditions to show greater gains than children in the Free Play condition showed. However, we have not yet tested the precise role of the playful context for vocabulary review. It could be the case that children showed better vocabulary gains in the adult-supported play conditions simply because the adults ensured additional vocabulary review in those conditions; the playful context for that review might not have been important. If the playful learning approach is uniquely effective, then children should learn words reviewed in a play context better than they learn words reviewed in an engaging context that also has adult support but uses more didactic methods. Study 2 was designed to address these issues. Also, in Study 2, our IS team trained classroom teachers to implement the activities themselves, allowing us to investigate whether teachers could adopt our program effectively. This shift increases the ecological validity of the research and the implications of our results.

The primary research questions in Study 2 were (1) Do children show improvements in vocabulary through a combination of book-reading and adult-supported play that are both delivered by classroom teachers? and (2) Do children learn more through the combination of book-reading and adult-supported playful learning sessions than they do through the combination of book-reading and a more direct teaching approach to supplementary vocabulary review that gives comparable exposure to the target words?

There were a variety of key differences in the designs of Study 1 and Study 2. For Study 2, in addition to the within-subjects comparison of pre- and post-test data, we used a within-subjects design to compare play and non-play review activities as supplements to the book-reading. There was no between-subjects condition comparison in Study 2. We also transitioned from having ISs deliver the intervention in small pull-out groups to having classroom teachers deliver the activities with IS support. Teachers ran the play sessions with small groups of participants but conducted book-readings with their full class. Further, in Study 2, we used only the

dragon theme, and classrooms ($N = 8$) were each assigned one of two books. This was more manageable for teachers, as they had the full two weeks to complete the activities for one book.

The number of target words per book was increased from 10 to 16, and we removed exposure words from the design to focus on the within-subjects comparisons between the two post-read review activities and between the target words and the 8 control words. For the play sessions in Study 2, since the Guided and Directed Play approaches were each effective in Study 1, we merged elements of both styles into the play approach for this study and kept the 10-min duration for each of the 4 play sessions per group. Words not reviewed in the play sessions were reviewed instead during a comparison activity: Teachers conducted 6 picture card review sessions with their full class, and these sessions tended to last up to 5 min each.

7. Method

7.1. Participants

We partnered with the same programs serving preschoolers from low-income families in Pennsylvania and Tennessee, and we followed similar IRB and recruitment procedures. Four classrooms from each state took part in the study, and all participating teachers were female. There were approximately 12 consented children per classroom ($N = 101$, 54.5% male; $M_{age} = 58.65$ months, $SD_{age} = 5.84$ months, $range_{age} = 40$–67 months). Teacher reports indicated that 13.9% of the children were classified as English Language Learners (ELL; $n = 14$). Parent reports confirmed that Spanish was spoken in the homes of most of those children and also in the homes of many children not designated as ELL. An additional 6.9% ($n = 7$) were missing both teacher and parent report data. Race and ethnicity information was not collected at the Tennessee site ($n = 52$) due to logistical constraints but was provided through parent questionnaires for 34 of the 49 children at the Pennsylvania site (20.6% African-American; 44.1% Hispanic; 35.3% Caucasian). Highest level of maternal education was reported by parent report for 84 children: 10.7% some high school, 40.5% high school diploma or GED, 13.1% trade school, 15.5% some college, 14.3% associate's degree, 1.2% bachelor's degree, and 4.8% graduate degree. Other background data are available in Tables A1 and A2. Comparisons across the sites showed similar patterns as those seen in Study 1, and moderation of condition effects by site was again examined.

7.2. Materials, intervention, and measures

7.2.1. Book and vocabulary word selection

For each of the two dragon-themed books, words were selected in a manner similar to Study 1, but with a focus only on nouns and verbs in the interest of simplifying an already complex design. Words from the original or revised text were used, and new words were added, as needed, to reach 16 target and 4 control words per book (see Table A5). Target words were split into sets A and B, with 8 words in each set and only one set being emphasized during a given reading or play activity to minimize overload. Given the within-subjects design for addressing question 2, Sets A and B were further divided into 4 Read + Play words (i.e., taught during reading and play) and 4 Read + Picture Card words (i.e., taught during reading and picture card review). Therefore, all children were taught all 16 words in the context of book reading, and they were also taught half those words during play and the other half during the picture card review activity. To facilitate comparisons of vocabulary gains across groups of words in this design, Read + Play words, Read + Picture Card words, and Control words in each book had equal distributions of abstract nouns, concrete nouns, and verbs.

Table 4
Study 2: word reviews per condition in within-subjects design.

	Read + play	Read + picture card
Number of words Reading sessions 1–4	8 (4 Set A and 4 Set B) (6 reviews) Multi-strategy review in 2 pre-readings and 2 readings (Set A Readings 1 & 3 or Set B Readings 2 & 4) and briefly defined during other 2 readings	8 (4 Set A and 4 Set B) (6 reviews) Multi-strategy review in 2 pre-readings and 2 readings (Set A Readings 1 & 3 or Set B Readings 2 & 4) and briefly defined during other 2 readings
Play sessions 1–4	(6 reviews) **Days 1 & 2:** multi-strategy review during 1 pre-play and 1 play session (Set A Day 1 or Set B Day 2) and briefly defined during 1 pre-play (the other Day) **Days 3 & 4:** used meaningfully in 2 pre-plays and given multi-strategy review in 1 play session (Set A Day 3 or Set B Day 4)	
Picture card sessions 1–6		(6 reviews) Multi-strategy review during each session
Total reviews	12	12

Note. Nature of vocabulary review for 16 target words over two weeks. All word sets (A and B of Read + Play words; and A and B of Read + Picture Card words) were reviewed 12 times: 6 times during reading sessions and 6 times during either play sessions or picture card sessions. Reviews consisted of either a brief definition, use of the word in a meaningful context, or multiple strategies (e.g., definitions, gestures, pictures), as outlined in the guidance materials for teachers.

7.2.2. Intervention procedure

Data were collected in April 2013. Although the teachers and students had been part of an interim phase of the Read-Play-Learn project (between Study 1 and Study 2) that involved the same books, each teacher was assigned to whichever book had not been used in her classroom during that phase. Therefore, while participants already had some exposure to the dragon theme, the story and words for this study had not been presented in her classroom previously. The two-week intervention consisted of multiple activities, all led by the classroom teacher: 4 book-reading sessions that reviewed Read + Play and Read + Picture Card words with the whole class (i.e., 12 participants plus up to approximately 8 other students); 4 small group play sessions that reviewed only the Read + Play words with each group of 3–4 study participants; and 6 whole-class picture card sessions that reviewed only the Read + Picture Card words (see Tables 4 and 5 and detailed descriptions below for information on equivalence and differences among reviews across condition). To minimize the level of project interference with regular classroom functioning and maximize implications for broader application, we intentionally chose to allow for natural variability in some elements of intervention delivery rather than to tightly control all elements. In that vein, teachers scheduled these activities based on their classroom needs, in consultation with the IS who was coaching them. In general, after a given reading and before the next one, teachers typically conducted a play session with each small group, such that each child experienced readings and play sessions in an alternating pattern. Picture card review sessions were also spread across the two weeks, interspersed among the other activities, at a time when the full class could be present and attentive.

Each teacher's assigned IS observed her reading and play sessions, contributed in-session prompts or clarifications (if desired), and provided brief feedback afterwards. Based on teacher report, average participant attendance was 3.75 of the 4 reading ses-

Table 5
Study 2: examples of word reviews.

Type of session	Type of review	Example
Pre-reading for readings 1–4	Multi-strategy review	"The dragon used his *nostrils* to blow fire and scare the knight. *Nostrils* are little holes in your nose. [Show picture card as you say definition.] Can you say that word? [Kids repeat.] Can you point to your nose? [Point to your own nose]. Do you have *nostrils*, too? Yes! You use them to breathe air, not fire!"
During readings 1–4	Multi-strategy review	"*Nostrils* are little holes in your nose that you use to breathe. [Point to picture as you say definition.] The dragon blows fire out of his *nostrils*, but we breathe air out of ours [point to your nostrils]. Can you say *nostrils* [kids repeat] and point to yours? [kids point]"
During readings 1–4	Brief definition	"*Nostrils* are little holes in your nose that you use to breathe. [Point to picture as you say this.]"
Pre-play for days 1–2	Multi-strategy review	"He is angry at the dragon. He bangs his foot down HARD to show he's angry, doesn't he? What do we call that? Yes, he is *stamping* his foot. Can you say *stamping*? [Children say word.] Show me how you do it. [Children do it.]"
Pre-play for days 1–2	Brief definition	"He is angry at the dragon, so he is *stamping* his foot. That means he is banging his foot down hard."
During play days 1–2	Multi-strategy review	[Incorporate word review based on children's play, at least 3 times.] "The egg is cracking! Something is *emerging* from the egg." "What is it called when someone comes out of something?"
Pre-play for days 3–4	Meaningful use (during toy distribution)	[Handing out the dragon figurine] "Don't forget about the fire that comes out of his *nostrils*!"
During play days 3–4	Multi-strategy review (incorporate based on children's play)	Use each of three review strategies: (1) a definition (e.g., "What *mayhem*! That is when there is a lot of mess and trouble."), (2) a closed-ended question (e.g., "Is *mayhem* calm or a little crazy?"), and (3) an open-ended question (e.g., "How is *mayhem* different than calm and peaceful?").
Picture card reviews 1–6	Multi-strategy review	"*Mayhem* is when there is a lot of trouble happening." [On later days, try having children guess the word when you provide the definition, before you show the picture] "Can you show me what that looks like?" [Gesture; kids repeat gesture.]

sions (*SD* = 0.55, range = 2–4 days) and 3.78 of the 4 play sessions (*SD* = 0.60, range = 1–4 days). One classroom teacher did not provide attendance data for her students (*n* = 12).

7.2.2.1. Book-reading. Teachers taught all 16 target words over the course of 4 reading sessions (see Table 4). Set A was the focus during sessions 1 and 3 and Set B was the focus during sessions 2 and 4. The words in the focus set were first reviewed in a pre-reading portion in which each of the day's 8 target words was taught through a scripted multistrategy review, including definitions, references to the word's use in the story, a picture card showing the word in a different context from the story, having children guess or repeat the word, and having children use gestures (see Table 5 for an example).

As in Study 1, during the book-reading portion of the session, the teacher used a script that outlined the use of similar strategies for reviewing the day's focus words (i.e., Set A or B) when they were encountered in the book's text. The script also explained how to briefly define the other set of words. For examples, see Table 5. Story comprehension prompts related to research questions not included in this paper were also incorporated into the script.

7.2.2.2. Play. As in Study 1, one complete set of book-related toys was available to the groups of children during 4 small-group play sessions approximately 10 min long. Guidance cards (similar to those used by ISs in Study 1's Guided Play condition) provided teachers with suggestions for the various portions of different play days, as well as general tips. The teacher was instructed to join children's play during all sessions and incorporate the day's designated target words into the play as smoothly as possible, according to general guidelines (see Table 4).

7.2.2.3. Days 1 & 2. Unlike in Study 1, play sessions did not usually follow immediately after book-readings. Therefore, the teacher started play Days 1 and 2 with a pre-play vocabulary and story review using illustrations from the book. On a given day, half of the Read + Play words were the main focus of the pre-play review (e.g.,

the 4 Read + Play words from Set A), and the other half were only briefly defined during this pre-play portion (e.g., the 4 Read + Play words from Set B). When the word was a focus word, the teacher was instructed to use multiple strategies to elaborate on the word's meaning, prompt children to say the word, and try to use gesture to illustrate meaning. On Day 2, when the word was not a focus word, the pre-play review was just a brief definition (see Table 5 for examples).

The play activity itself incorporated some of the adult-provided re-enactment structure that was integral to the Directed Play condition in Study 1. However, while the ISs in Study 1 used a script to deliver precise directions to children, here teachers were asked to gently suggest ideas through questions and use of the toys while respecting children's choices, as in the Guided Play condition in Study 1. The guidance cards suggested how to playfully relate target words to what children might be doing and provided brief definitions and example questions to prompt children's word usage (see Table 5). The goal was to incorporate each of the day's 4 assigned words (i.e., from Set A on Day 1, Set B on Day 2) at least three times, using multiple strategies.

7.2.2.4. Days 3 & 4. On Days 3 and 4, the teacher began by presenting children with a choice of story re-enactment or unrelated play scenarios (i.e., beach, birthday party). Offering children choices was a way to give them control, akin to the Guided Play condition in Study 1, but also provide some structure, which might have been a beneficial component to the Directed Play condition of Study 1. There were guidance cards specific to each scenario. The pre-play portion for these days did not include story review, since the children had already encountered the story repeatedly by that point. Instead, the guidance card had suggestions for briefly using all 8 Read + Play words in a meaningful way during toy distribution. Then, there were suggestions for the teacher for helping children begin to play (e.g., "How is everyone going to get to the playground?"). These days were thus similar to Study 1's Guided Play condition, but with suggestions provided to teachers for pos-

sible scenes that they might use to help guide children in a less formal manner than was used in Study 1's Directed Play condition. Teachers were instructed to use each of the day's 4 target vocabulary words (i.e., from Set A on Day 3, Set B on Day 4) at least once during play with each of three vocabulary review strategies: (1) a definition, (2) a closed-ended question, and (3) an open-ended question. Teachers could use examples from the guidance card (see Table 5) or develop their own.

7.2.2.5. Picture card review. At three convenient times per week, the teacher spent about 5 min reviewing the 8 words in the Read + Picture Card condition using picture cards with illustrations different from those in the book. They used similar methods employed during book readings, including brief definitions, gestures, and prompts for children to say the words (see Table 5 for examples). In order to have each Read + Picture card word reviewed during this activity the same number of times (6) as each of the Read + Play words were reviewed during play, this review occurred 6 times (outside of the book-reading sessions), on 6 days within the two-week period.

7.3. Measures

Children were pre- and post-tested on both the receptive and expressive measures, which included the 4 control words and 16 target words associated with the book they were read. On average, 27.74 days (SD = 8.74) elapsed between pre- and post-testing. Time elapsed was not significantly related to post-test scores. Scoring and coding procedures were the same as in Study 1, with internal consistency for the receptive measure of Cronbach's α = .57 and .69 for the two books (*Dragon for Breakfast* and *Knight and Dragon*, respectively), and an average percent agreement for the 20% of cases double scored for reliability (between 3 masters' level coders and the Gold Standard coder) on the expressive task of 96.3% (Cohen's kappa = 0.93). (See Study 1 Measures section for a more detailed interpretation of all measures' procedures and statistical qualities. One point of note is that in Study 2, there were no words that were unable to be depicted on the receptive test.) As in Study 1, parent questionnaires were sent home with children, and 89 were returned with all or most items answered.

Additionally, the majority of participants (n = 97) were pretested on the Peabody Picture Vocabulary Test (PPVT-IV; Dunn & Dunn, 2007) approximately 4 weeks prior to the first week of the intervention activities of Study 2. Participants were also tested on measures of narrative retelling and narrative comprehension, which are not reported on in this manuscript.

7.4. Fidelity of implementation

In addition to having ISs observe the book-reading and play sessions and provide feedback and support to teachers, we also documented teachers' fidelity of implementation by noting the proportion of assigned words that teachers reviewed. For book-reading, ISs monitored word reviews during at least one of each teacher's first two sessions; every teacher reviewed all 8 assigned words. For play sessions, ISs monitored words reviewed during all 4 days of all play-groups for each teacher. The lowest score for a single session was 75%, which reflects a teacher missing 1 out of 4 assigned words for the day, and no teacher received that score more than once. Every teacher averaged at least 97%.

7.5. Data analyses

We used MLM to test if children's pre- to post-intervention residualized vocabulary gains were different for words taught by Read + Play, words taught by Read + Picture Cards, and non-taught

Table 6
Study 2: Receptive and expressive vocabulary unadjusted means (standard deviations) by assessment session and condition.

Measures	Pre-test Mean (SD)	Post-test Mean (SD)
Read + Play		
Receptive vocabulary	0.38 (0.17)	0.70 (0.20)**
Expressive vocabulary	0.09 (0.53)	0.85 (0.14)**
Read + Picture cards		
Receptive vocabulary	0.43 (0.17)	0.70 (0.22)**
Expressive vocabulary	0.08 (0.47)	0.64 (0.11)**
Control		
Receptive vocabulary	0.37 (0.23)	0.45 (0.22)*
Expressive vocabulary	0.08 (0.33)	0.17 (0.23)*

Note. N = 101. Receptive task values indicate the proportion of items that were answered correctly, with chance level equal to 0.33 (i.e., select one of three images). Expressive task values indicate the average score per expressive item, and these scores ranged from 0.00 to 2.50. For comparison, the range for Hadley et al. (2016) was 0–2.02.
Pre- to post-test gains significant at ** p < .01, * p < .05.

Control words, which were repeated within children (i.e., children provided data for all three conditions). MLM also accounted for the nesting of children within classrooms. For a given outcome, analyses were conducted on all available cases, with fewer than 3% of children having missing data on one of the two vocabulary tests.

8. Results

Table 6 shows mean vocabulary scores and standard deviations by time and condition. Correlations between pre-test PPVT scores and pre-test receptive and expressive scores on the Read + Play and Read + Picture Card words were positive and significant (ranging from r = .21 to r = .35), which contributes to the construct validity of our project-specific measures.

8.1. Children's vocabulary gains over time

Regarding our first research question, a model similar to Equation 1 revealed that children had significantly (ps < .01) larger receptive and expressive scores at post-test compared to pre-test for Read + Play words ($B_{receptive}$ = 0.31, SE = 0.02, d = 1.66; $B_{expressive}$ = 0.76, SE = 0.05, d = 1.95) and Read + Picture Card words ($B_{receptive}$ = 0.28, SE = 0.02, d = 1.42; $B_{expressive}$ = 0.56, SE = 0.04, d = 1.63). They also had small but statistically significant gains (ps < .05) for control words ($B_{receptive}$ = 0.08, SE = 0.03, d = 0.34; $B_{expressive}$ = 0.09, SE = 0.04, d = 0.31).

8.2. Comparison of play versus picture card review conditions

To address our second research question, we compared the gains for Read + Play words and Read + Picture Card words, as well as for Control words, utilizing a within-subjects design (see Table 7). We examined children's gains by controlling for pre-test scores, even though tests of baseline equivalence showed no significant differences (ps > .10) on pre-test receptive or expressive scores among Read + Play, Read + Picture Card, and control words. Our model accounted for level of instruction$_i$ nested within children$_j$, nested within classroom$_k$ (see Supplementary Materials, Eq. (4)). Analysis indicated that children showed greater gains for Read + Play words than for Control words ($B_{receptive}$ = −0.16, SE = 0.03, p < .01, d = 1.18; $B_{expressive}$ = −0.50, SE = 0.04, p < .01, d = 1.55). They also showed greater gains for Read + Picture Card words than for Control words ($B_{receptive}$ = −0.17, SE = 0.03, p < .01, d = 1.12; $B_{expressive}$ = −0.41, SE = 0.04, p < .01, d = 1.18). Although children showed no greater gains in receptive knowledge of Read + Play words over Read + Picture Card words (B = 0.01, SE = 0.03, p > .05,

Table 7

Study 2: Fixed effect estimates (standard errors) for models estimating differences in level of instruction for receptive and expressive vocabulary gains at post-test (play = reference group).

Parameters	Receptive	Expressive
Intercept, γ_{000}	0.492 (0.038)	0.588 (0.062)
Pre-test performance, γ_{100}	0.386 (0.067)[**]	0.747 (0.164)[**]
Read + Play versus read + picture cards, γ_{200}	0.007 (0.029)	−0.090 (0.041)[*]
Read + Play versus control, γ_{300}	−0.163 (0.029)[**]	−0.497 (0.041)[**]
Read + Picture cards versus control	−0.170 (0.029)[**]	−0.408 (0.041)[**]

Note. N = 101. Standard errors adjusted for interdependency among observations. Level of instruction (Level 1) was repeated within children (Level 2) who were nested within classrooms (Level 3). Read + play words are the reference group (γ_{200}, γ_{300}) for the comparison of level of instruction difference; as such, negative estimates indicate that children made greater gains on Read + Play words compared to the other types of words. Contrast of Read Only (reference group) versus Control words was estimated with post-hoc pairwise comparison.

[*] *p* < .05.
[**] *p* < .01.

d = 0.01), they showed significantly greater expressive gains for Read + Play words than for Read + Picture Card words (*B* = −0.09, *SE* = 0.04, *p* < .05, *d* = 0.41).

We also investigated whether any of the following factors moderated the comparison between Read + Play words and Read + Picture Card words: children's composite home literacy exposure, dichotomized maternal education, composite home language usage, ELL status, pre-test scores on the target words, pre-test PPVT scores, attendance, and site. There was no statistically significant moderation of the comparison between those two conditions.

9. Discussion of study 2

In answer to our first research question, children successfully learned new words, whether the words were taught through book-reading and play or through book-reading and the picture card activity. Their gains for these sets of target words were greater than their gains for control words, suggesting that they learned through the program and not just through general development. As for the second research question, for the depth of vocabulary knowledge tested on the expressive task, it was more effective to review words in the context of the play activity than in the context of the more direct teaching picture card activity, despite equal numbers of word reviews (6 outside of book-reading).

It could be the case that the play activity was more effective because it included adults' invitations for children to both use the words and reflect on their meaning. In that vein, preliminary additional analyses of Study 1 data suggest that children's active engagement in answering open- and closed-ended questions about words during play was associated with greater knowledge gains for certain types of words, even after controlling for frequency of the adult's word usage (Ilgaz, Weisberg, Hirsh-Pasek, Golinkoff, & Nicolopoulou, 2013). There is also other research showing that teachers' question-asking during play is associated with children's verbal responsiveness, especially when the questions are open-ended or build on children's contributions (de Rivera, Girolametto, Greenberg, & Weitzman, 2005; Meacham, Vukelich, Han, & Buell, 2014). Although both our play and our picture card activity incorporated elements of testing children's knowledge, such as by challenging children to recall words themselves, the play activity invited children to engage with words more meaningfully through answering questions.

While we found a moderate-sized effect favoring the Read + Play condition over the Read + Picture Cards condition for the expressive measure, there was no significant difference on our receptive measure. One possibility is that word learning is a two-stage process in which initial fast-mapped representations can be formed relatively quickly, while deeper learning, such as is tapped by our expressive measure, requires more active processing. Deeper processing may require the establishment of connections to long-term semantic knowledge that is needed to respond accurately to our expressive measure. Support for this speculation comes from an fMRI study that revealed changes in neural activity after a word has been consolidated overnight (Davis, Di Betta, Macdonald, & Gaskell, 2009).

Overall, children's gains in knowledge of taught words in Study 2 were even greater than gains seen in Study 1. This difference could be related to their being taught and tested on fewer words total. Further research should more precisely address the impacts of word distribution and total number of words (see also Wasik et al., 2016). In conjunction with the shift in delivery from ISs to teachers, these results bode well for one of the ultimate aims of our project: to create an intervention that preschool teachers can implement effectively.

10. General discussion

In both Studies 1 and 2, children showed statistically significant increases in vocabulary knowledge after experiencing the Read-Play-Learn intervention, and gains were greatest for words actively taught. Furthermore, growth was apparent on both receptive and expressive measures, which capture different types of knowledge.

10.1. Play with adult support can promote vocabulary growth

In line with prior work examining various learning goals (Chien et al., 2010; Fisher et al., 2013; Klahr & Nigam, 2004), Study 1 showed that adult-supported play with story-related toys is more effective for vocabulary growth than children's free play with those toys. When adults actively incorporated vocabulary review in play, children heard the adult model usage of the words within the play context, define the words again, and discuss the words' meanings through interactive questioning. Through this adult-supported play, children learned more than they did without additional review beyond the book-reading activity and unstructured exposure to related toys (Study 1's Free Play condition). This play-based context illustrates that learning takes place best when there is frequency of exposure to the material and diversity of context within a meaningful and socially interactive environment that invites children to be active and engaged (Harris et al., 2011; Meltzoff et al., 2009). Although in Study 1 the comparative benefits of different approaches to adult support depended on maternal education, the lack of a maternal education moderator in Study 2 indicates that our new play approach, which progressed from more directed to more guided, helped children regardless of that background.

10.2. Why does adult-supported play promote vocabulary growth?

A key question that remained after Study 1 was whether play with adult support was effective due to play *per se* or whether supplementary vocabulary review strategies presented in a non-play context would be equally effective. In Study 2, both play and picture cards were equally effective review methods for establishing initial representations, as measured by our receptive task. However, Study 2 showed that learning through adult-supported play bolstered children's deeper (expressive) knowledge of words compared to the picture card activity. This suggests unique benefits of review within the play context, beyond the similar numbers of exposures that both interactive activities incorporated. The effectiveness of guided play in promoting children's learning might be

attributed to the positive emotion and self-directed *mise en place* that play creates for children (Weisberg, Hirsh-Pasek, Golinkoff, & McCandliss, 2014). Because children were actively involved in constructing the play scenario, this activity was more meaningfully connected to children's interests and hence inherently more conducive to deeper engagement with the vocabulary words.

10.3. Implications for education policy and practice

As debates continue about how to design preschool programming to best prepare children for kindergarten and beyond (Fuller et al., 2017; Zigler, Gilliam, & Barnett, 2011), our findings point to both book-reading and play as promising approaches. Children learned words during our book-reading sessions, even without supplementary activities explicitly incorporating word review (i.e., in the Free Play condition in Study 1). Vocabulary gains were further boosted with the addition of adult-supported review (i.e., Guided Play or Directed Play in Study 1; Play or Picture Cards in Study 2). Last, the play-based approach was an especially effective supplementary review activity for boosting expressive knowledge (i.e., Play vs. Picture Cards in Study 2).

Effect sizes associated with the differences in children's gains for target words versus for control words were medium (0.58 for Study 1 receptive) to large (0.89 for Study 1 expressive), with Study 2 results showing especially large effects (between 1.12 and 1.55 for both tasks). These large effect sizes are particularly compelling given that most studies of vocabulary interventions using book-reading report effect sizes closer to 0.45 or 0.66 (Mol et al., 2009; National Early Literacy Panel, 2008). In addition, our large Study 2 effects were achieved through teacher delivery and were greater than the effects we found in an interim phase of our project with similar methods but delivery by ISs.

The contribution of play towards vocabulary learning in our work indicates that trends to reduce play in favor of activities thought to better support academic goals (Elkind, 2008; Miller & Almon, 2009) may be misguided regarding vocabulary development, as well as other domains (Fisher et al., 2010). Children benefitted from the play-based approaches, suggesting that the pursuit of similar strategies in our classrooms is a promising option for supporting word-learning. The benefits of academic-oriented programming (Fuller et al., 2017) can be achieved within a playful context.

10.4. Limitations and future directions

While the larger effect sizes obtained through teacher implementation versus IS implementation is encouraging, several factors varied between the studies, adding caveats to any comparisons. These designs do not allow us to determine the role of conducting book-reading with a small group (Study 1) versus the whole class (Study 2) (but see Neuman & Kaefer, 2013), or whether children's learning differs when reading and play are experienced back-to-back (Study 1) versus on different days (Study 2). Similar questions arise regarding child characteristics or elements of home environments that could impact effectiveness. Although we tested for many moderation effects, there might be additional elements, such as quality of home language input that we did not sufficiently capture in our questionnaire but could incorporate into future work.

Similarly, it will be important to see if the effects of our methods generalize to other contexts beyond the Head Start and pre-K programs participating in these studies. It is possible that characteristics of these programs, such as their curricula or the populations they serve, contributed to the results in ways that were not addressed by our research questions. Another related unanswered question is about how much coaching classroom teachers need to help them achieve fidelity and effectiveness. ISs observed and gave feedback to teachers on their read and play sessions in Study 2, but we did not examine the role of that coaching in detail. In addition, our methods were delivered using carefully selected books, vocabulary words, and prepared guidance material. We do not yet know what would happen if we ask teachers to use similar techniques without materials that are as specifically tailored and prepared for this purpose.

Nor do we yet know which elements of our activities are the key ingredients responsible for vocabulary gains. Identification of these essential components would inform efforts in practical settings. These two studies attempted to probe the added value of play, when combined with more well-studied book-reading methods. As such, we held book-reading constant across conditions in each study because we were focused on the role of play in providing vocabulary support beyond what is provided in the book-reading approach; however, it would be useful moving forward to understand play's power to facilitate vocabulary development without the initial book-reading component, and we are pursuing that question in further studies. It will also be important to determine if similar effects are attainable the first time students and teachers experience these approaches or whether some of the gains we observed may have been due to participants' prior experiences in our project. It is possible that familiarity with the methods assists with learning, even when working with new words, books, and play props.

Prior to efforts toward generalization and dissemination of our methods, it is important to consider potential alternative explanations regarding the comparison of conditions in our studies. To isolate the effect of the different pedagogical styles, we selected aspects of play and, later, of picture card activities to be kept constant across conditions. However, other differences in the conditions besides the context and delivery of review might have led to the disparate gains across conditions. For example, the first two adult-supported play sessions in Study 2 included pre-play vocabulary and story review to refresh children's memories from the book-reading sessions that happened on earlier days. In contrast, the Read + Picture Card words were not reviewed before the activity. Another difference between the conditions was that the play was conducted in small groups while the picture card review was with the whole class; however, prior literature (Marulis & Neuman, 2010; Neuman & Kaefer, 2013) suggests that the size of the group is unlikely to account for the observed differences in learning. Additionally, teachers' scheduling of the various activities in Study 2 may have led to differences in the time of day when play versus picture card review took place or the amount of time between the readings and these activities. Although such differences were intentionally allowed by our choice to embrace natural variability, the differences may be meaningfully systematic and may have impacted children's learning from each activity.

There was also a difference in the total amount of time spent in play (approximately 10 min for 4 sessions = 40 min) versus in the picture card review activity (approximately 5 min for 6 sessions = 30 min), although these time periods reflect the ecological validity of what happens in these situations naturally within a classroom. Furthermore, those totals do not reflect the duration of time spent on reviewing the words. Some of the time spent in play was not devoted to vocabulary, since the words were embedded within the larger task of using the toys to enact a scene. In contrast, the picture card activity time was more focused on vocabulary. Rather than equating the time spent on reviewing words, which is harder to control, we opted to aim for a balance in the number of times (6) each individual word was reviewed within play or within the picture card review. Future work could address whether the distribution of those 6 instances across 4 play sessions versus 6 picture card review sessions might have played any role in the effectiveness of the activities.

Additionally, "reviewing" a word in the different activities involved one or more of a variety of behaviors, such as using the word in a sentence, defining the word, asking children to say the word, using a gesture, asking children to gesture, and asking closed- and open-ended questions about the word. Although we kept numerical equivalency in mind, differences in what "review" means within the play and picture card activities are inherent to the core of the pedagogical approaches themselves. Asking children questions about words is in some ways more natural and meaningful in the play context, whereas having teachers ask children questions during the picture card review might have slowed down the activity and made it more forced and less enjoyable. Our main goal was to design the picture card activity to provide children with a comparable and fun opportunity to review and recall vocabulary words and their meanings and to apply their knowledge in a context beyond that of the storybook itself. It remains difficult, however, to tease apart the effects of play *per se* from the effects of the type of review to which the play context lends itself. In terms of basic research and efforts to understand underlying mechanisms for vocabulary learning, these issues invite further inquiry.

Last, we did not examine the effects of our methods on other aspects of language development or on skills with previously established bidirectional associations with language learning, such as executive functioning (Fuhs, Nesbitt, Farran, & Dong, 2014). Future research could expand the investigation of these reading- and play-based pedagogical approaches for promoting various areas of social, emotional, and cognitive development in young children with low socioeconomic backgrounds and in other populations.

10.5. Conclusions

Here, we studied methods for promoting low-income children's vocabulary learning, directly comparing different types of play activities as supplements shared book-reading. We also compared play to a word review activity that used a more direct teaching approach. We found that adult-supported play is especially promising for promoting preschoolers' deeper word-learning. Results suggest there *is* a role for play in our early education classrooms—particularly in low-income communities, which paradoxically encounter the strongest pressures to limit play to focus on improving academic outcomes. While others have argued that "All work and no play makes Jack a dull boy" (Howell, 1659) and "All play and no work makes Jack a mere toy" (Edgeworth, 1825, p. 155), we add through our work: Learning through play makes Jack's learning a joy.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ecresq.2018.01.010.

References

Asaridou, S. S., Demir-Lira, E., Goldin-Meadow, S., & Small, S. L. (2017). The pace of vocabulary growth during preschool predicts cortical structure at school age. *Neuropsychologia*, 98, 13–23. http://dx.doi.org/10.1016/j.neuropsychologia.2016.05.018

Ash, D., & Wells, G. (2006). Dialogic inquiry in classroom and museum: Action, tools, and talk. In Z. Bekerman, N. C. Burbules, & D. Silberman-Keller (Eds.), *Learning in places: The informal education reader* (Vol. 249) (pp. 35–54).

Beck, I. L., & McKeown, M. G. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, 107(3), 251–271. http://dx.doi.org/10.1086/511706

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). New York: The Guilford Press.

Biemiller, A. (2010). *Words worth teaching: Closing the vocabulary gap*. Columbus, OH: SRA/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. http://dx.doi.org/10.1037/0022-0663.98.1.44

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. http://dx.doi.org/10.1037/a0013844

Bracken, S. S., & Fischel, J. E. (2008). Family reading behavior and early literacy skills in preschool children from low-income backgrounds. *Early Education and Development*, 19(1), 45–67. http://dx.doi.org/10.1080/10409280701838835

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences USA*, 110(28), 11278–11283. http://dx.doi.org/10.1073/pnas.1309518110

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale–Chall readability formula*. Cambridge, MA: Brookline Books/Lumen Editions.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105. http://dx.doi.org/10.1111/j.1756-8765.2008.01005.x

Chien, N. C., Howes, C., Burchinal, M., Pianta, R. C., Ritchie, S., Bryant, D. M., . . . & Barbarin, O. A. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development*, 81(5), 1534–1549.

Christ, T., & Wang, X. C. (2011). Closing the vocabulary gap? A review of research on early childhood vocabulary practices. *Reading Psychology*, 32(5), 426–458. http://dx.doi.org/10.1080/02702711.2010.495638

Cooper, H. (2005). *Pumpkin soup*. New York: Square Fish.

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.

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

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, 21(4), 803–820. http://dx.doi.org/10.1162/jocn.2009.21059

dePaola, T. (1998). *The knight and the dragon*. New York: The Putnam & Grosset Group.

de Rivera, C., Girolametto, L., Greenberg, J., & Weitzman, E. (2005). Children's responses to educators' questions in day care play groups. *American Journal of Speech-Language Pathology*, 14(1), 14–26.

Dickinson, D. K., Hofer, K. G., Barnes, E. M., & Grifenhagen, J. F. (2014). Examining teachers' language in head start classrooms from a systemic linguistics approach. *Early Childhood Research Quarterly*, 29(3), 231–244. http://dx.doi.org/10.1016/j.ecresq.2014.02.006

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: Preschool language experiences and later language and reading. *Child Development*, 82(3), 870–886. http://dx.doi.org/10.1111/j.1467-8624.2011.01576.x

Dickinson, D.K., Nesbitt, K.T., Collins, M. F., Hadley, E. B., Newman, K., Rivera, B., Ilgez, H., Nicolopoulou, A., Golinkoff, R. M., & Hirsh-Pasek-R. (in preparation). Using book reading to teach for breath and depth of vocabulary knowledge.

Dollaghan, C. A., Campbell, T. F., Paradise, J. L., Feldman, H. M., Janosky, J. E., Pitcairn, D. N., & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research*, 42(6), 1432–1443. http://dx.doi.org/10.1044/jslhr.4206.1432

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test (PPVT-IV)*. San Antonio, TX: Pearson Clinical Assessments.

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D.-M. C., Crawford, G. M., . . . & Pianta, R. C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, 25(2), 177–193. http://dx.doi.org/10.1016/j.ecresq.2009.10.003

Edgeworth, M. (1825).. *Harry and Lucy concluded: Being the last part of early lessons* (Vol. 2) London: Munroe & Francis. Retrieved from https://archive.org/details/harryandlucycon01unkngoog

Elkind, D. (2008). . pp. 14–17. *Can we play?* (vol. 4) Greater Good Magazine.

Farkas, G., & Beron, K. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Science Research*, 33(3), 464–497. http://dx.doi.org/10.1016/j.ssresearch.2003.08.001

Fisher, K., Hirsh-Pasek, K., Golinkoff, R. M., Singer, D. G., & Berk, L. (2010). Playing around in school: Implications for learning and educational policy. In P. Nathan, & A. D. Pellegrini (Eds.), *The Oxford handbook of the development of play*. Retrieved from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780195393002.001.0001/oxfordhb-9780195393002-e-025

Fisher, K., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. *Child Development*, 84(6), 1872–1878. http://dx.doi.org/10.1111/cdev.12091

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

Fuller, B., Bein, E., Bridges, M., Kim, Y., & Rabe-Hesketh, S. (2017). Do academic preschools yield stronger benefits? Cognitive emphasis, dosage, and early learning. *Journal of Applied Developmental Psychology*, 52, 1–11. http://dx.doi.org/10.1016/j.appdev.2017.05.001

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology*, 64, 257–283. http://dx.doi.org/10.1146/annurev-psych-113011-143802

Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., & Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study: Hypotheses for intervention. *American Psychologist*, 69(6), 588–599. http://dx.doi.org/10.1037/a0036886

Greenberg, M. T., & Abenavoli, R. (2017). Universal interventions: Fully exploring their impacts and potential to produce population-level impacts. *Journal of Research on Educational Effectiveness*, 10(1), 40–67. http://dx.doi.org/10.1080/19345747.2016.1246632

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.

Han, M., Moore, N., Vukelich, C., & Buell, M. (2010). Does play make a difference? How play intervention affects the vocabulary learning of at-risk preschoolers. *American Journal of Play*, 3(1), 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). NY: Guilford Press.

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

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 gameplay. *Mind, Brain, and Education*, 10, 71–80. http://dx.doi.org/10.1111/mbe.12103

Hassinger-Das, B., Toub, T. S., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). A matter of principle: Applying language science to the classroom and beyond. *Translational Issues in Psychological Science*, 3(1), 5–18. http://dx.doi.org/10.1037/tps0000085

Hindman, A. H., & Wasik, B. A. (2013). Vocabulary learning in Head Start: Nature and extent of classroom instruction and its contributions to children's learning. *Journal of School Psychology*, 51(3), 387–405. http://dx.doi.org/10.1016/j.jsp.2013.01.001

Hindman, A. H., Wasik, B. A., & Snell, E. K. (2016). Closing the 30 million word gap: Next steps in designing research to inform practice. *Child Development Perspectives*, 10(2), 134–139. http://dx.doi.org/10.1111/cdep.12177

Hirsh-Pasek, K., & Golinkoff, R. M. (2011). The great balancing act: optimizing core curricula through playful learning. In E. Zigler, W. S. Gilliam, & W. S. Barnett (Eds.), *The pre-K debates: Current controversies and issues* (pp. 110–115). Baltimore, Md: Brookes Publishing Company.

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

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, 26(7), 1071–1083. http://dx.doi.org/10.1177/0956797615581493

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in educational apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, 16(1), 3–34. http://dx.doi.org/10.1177/1529100615569721

Hoff, E. (2013). Interpreting the early language trajectories of children from low-ses and language minority homes: Implications for closing achievement gaps. *Developmental Psychology*, 49(1), 4–14. http://dx.doi.org/10.1037/a0027238

Howell, J. (1659). *Paroimiographia proverbs, or, old sayed savves & adages in English (or the Saxon toung) Italian, French, and Spanish, whereunto the British for their great antiquity and weight are added*. London: J. G. Retrieved from http://name.umdl.umich.edu/A44738.0001.001

Ilgaz, H., Weisberg, D., Hirsh-Pasek, K., Golinkoff, R. M., & Nicolopoulou, A. (2013, April). Not all play is created equal: When playful learning sparks vocabulary acquisition in low-income children. In D. Dickinson (Chair), Effects of varied types of adult-supported play on preschool children's receptive vocabulary learning. Society for Research on Child Development, Seattle, WA.

Jenkins, J. M., & Duncan, G. J. (2017). *Do pre-kindergarten curricula matter? (The current state of scientific knowledge on pre-kindergarten effects)*. Washington, DC: Brookings Institution and Duke Center for Child and Family Policy. Retrieved from https://sanford.duke.edu/articles/pre-kindergarten-effects-%E2%80%93-what-science-says

Johnson, J. E., Christie, J. F., & Yawkey, T. D. (1999). *Play and early childhood development*. New York, NY: Longman.

Justice, L. M., Mashburn, A. J., Hamre, B. K., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, 23(1), 51–68. http://dx.doi.org/10.1016/j.ecresq.2007.09.004

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15(10), 661–667.

Konishi, H., Kanero, J., Freeman, M. R., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Six principles of language development: Implications for second language learners. *Developmental Neuropsychology*, 39(5), 404–420. http://dx.doi.org/10.1080/87565641.2014.931961

Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin*, 139(1), 1–34. http://dx.doi.org/10.1037/a0029321

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.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., . . . & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732–749. http://dx.doi.org/10.1111/j.1467-8624.2008.01154.x

McMullen, E., & McMullen, N. (1990). *Dragon for Breakfast*. Minneapolis: Carolrhoda Books.

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

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, 325(5938), 284–288. http://dx.doi.org/10.1126/science.1175626

Miller, E., & Almon, J. (2009). *Crisis in the kindergarten: Why children need to play in school*. College Park, MD: Alliance for Childhood.

Miller, E. B., Farkas, G., Vandell, D. L., & Duncan, G. J. (2014). Do the effects of Head Start vary by parental preacademic stimulation? *Child Development*, 85(4), 1385–1400. http://dx.doi.org/10.1111/cdev.12233

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education and Development*, 19(1), 7–26. http://dx.doi.org/10.1080/10409280701838603

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. http://dx.doi.org/10.3102/0034654309332561

Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-Month-old children with larger oral vocabularies display greater academic and behavioral functioning at kindergarten entry. *Child Development*, 86(5), 1351–1370. http://dx.doi.org/10.1111/cdev.12398

National Early Literacy Panel. (2008). *Developing early literacy*.. Retrieved from http://www.lincs.ed.gov/publications/pdf/NELPReport09.pdf

Neuman, S. B., & Dwyer, J. (2009). Missing in action: Vocabulary instruction in pre-k. *The Reading Teacher*, 62(5), 384–392. http://dx.doi.org/10.1598/RT.62.5.2

Neuman, S. B., & Kaefer, T. (2013). Enhancing the intensity of vocabulary instruction for preschoolers at risk: The effects of group size on word knowledge and conceptual development. *The Elementary School Journal*, 113(4), 589–608. http://dx.doi.org/10.1086/669937

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.

Nicolopoulou, A., & Ilgaz, H. (2013). What do we know about pretend play and narrative development? A response to Lillard, Lerner, Hopkins, Dore, Smith, and Palmquist on the impact of pretend play on children's development: A review of the evidence. *American Journal of Play*, 6(1), 55.

Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., . . . & Weiland, C. (2017). *Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects ? A consensus statement*. Washington, DC: Brookings Institution and Duke Center for Child and Family Policy. Retrieved from https://sanford.duke.edu/articles/pre-kindergarten-effects-%E2%80%93-what-science-says

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. http://dx.doi.org/10.1080/10409289.2016.1220771

Qi, C. H., Kaiser, A. P., Milan, S., & Hancock, T. (2006). Language performance of low-income African American and European American preschool children on the PPVT-III. *Language, Speech, and Hearing Services in Schools*, 37(1), 5–16.

Reed, J., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Learning on hold: Cell phones sidetrack parent–child interactions. *Developmental Psychology*.

Richels, C. G., Johnson, K. N., Walden, T. A., & Conture, E. G. (2013). Socioeconomic status, parental education, vocabulary and language skills of children who

stutter. *Journal of Communication Disorders*, *46*(4), 361–374. http://dx.doi.org/10.1016/j.jcomdis.2013.07.002

Rogoff, B. (2003). *The cultural nature of human development.* Oxford: Oxford University Press.

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

Roskos, K. A., & Christie, J. F. (2000). *Play and literacy in early childhood: Research from multiple perspectives.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Roskos, K. A., Tabors, P. O., & Lenhart, L. A. (2004). *Oral language and early literacy in preschool: Talking, reading, and writing.* Newark, DE: International Reading Association, Inc.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1) http://dx.doi.org/10.1017/S0305000907008343

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development*, *83*(2), 508–525. http://dx.doi.org/10.1111/j.1467-8624.2011.01710.x

Rowe, M. L., Silverman, R. D., & Mullan, B. E. (2013). The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology*, *38*(2), 109–117. http://dx.doi.org/10.1016/j.cedpsych.2012.12.001

Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language*, *24*(1), 123–138. http://dx.doi.org/10.1017/S0305000996003005

Singer, D. G., & Singer, J. L. (1990). *The house of make-believe: Children's play and the developing imagination.* Cambridge, MA: Harvard University Press.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academies Press. Retrieved from http://www.nap.edu/catalog/6023

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, *38*(6), 934–947. http://dx.doi.org/10.1037//0012-1649.38.6.934

Toub, T. S., Rajan, V., Golinkoff, R. M., & Hirsh-Pasek, K. (2016). Guided play: A solution to the play versus learning dichotomy. In D. Geary, & D. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp. 117–141). Switzerland: Springer International Publishing. Retrieved from http://www.springer.com/us/book/9783319299846

Vernon-Feagans, L., Bratsch-Hines, M. E., & The Family Life Project Key Investigators. (2013). Caregiver–child verbal interactions in child care: A buffer against poor language outcomes when maternal language input is less. *Early Childhood Research Quarterly*, *28*(4), 858–873. http://dx.doi.org/10.1016/j.ecresq.2013.08.002

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

Waddell, M., & Oxenbury, H. (1991). *Farmer duck.* Somerville, MA: Candlewick 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. http://dx.doi.org/10.1037/0022-0663.98.1.63

Wasik, B. A., Hindman, A. H., & Snell, E. K. (2016). Book reading and vocabulary development: A systematic review. *Early Childhood Research Quarterly*, *37*(Suppl. C), 39–57. http://dx.doi.org/10.1016/j.ecresq.2016.04.003

Weisberg, D. S., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Guided play: Where curricular goals meet a playful pedagogy. *Mind, Brain, and Education*, *7*(2), 104–112. http://dx.doi.org/10.1111/mbe.12015

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*, http://dx.doi.org/10.1177/0963721416645512

Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., & McCandliss, B. D. (2014). Mise en place: Setting the stage for thought and action. *Trends in Cognitive Sciences*, *18*(6), 276–278. http://dx.doi.org/10.1016/j.tics.2014.02.012

Weisberg, D. S., Kittredge, A. K., Hirsh-Pasek, K., Golinkoff, R. M., & Klahr, D. (2015). Making play work for education. *Phi Delta Kappan*, *96*(8), 8–13. http://dx.doi.org/10.1177/0031721715583955

Weisberg, D. S., Zosh, J. M., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Talking it up: Play, language development, and the role of adult support. *American Journal of Play*, *6*(1), 39.

What Works Clearinghouse. (2014). *Procedures and standards handbook (No. 3.0).* Washington, DC: Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_handbook.pdf

Zigler, E., Gilliam, W. S., & Barnett, W. S. (Eds.). (2011). *The pre-K debates: Current controversies and issues.* Baltimore, MD: Brookes Publishing Company.