Early Childhood Research Quarterly 50 (2020) 101–113

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Increasing preschoolers' vocabulary development through a streamlined teacher professional development intervention☆

Barbara A. Wasik*, Annemarie H. Hindman

Temple University, United States

ARTICLE INFO

Article history:
Received 18 January 2018
Received in revised form 31 October 2018
Accepted 11 November 2018
Available online 21 December 2018

Keywords:
Vocabulary development
Preschool intervention
ECE teacher professional development
Vocabulary gap

ABSTRACT

Preschool teachers from a high-poverty, urban school district were trained to implement Story Talk, a book reading intervention designed to increase children's vocabulary and language development using supportive materials and strategic individualized coaching. Thirty-five teachers were randomly assigned by site to the intervention (20) or the control condition (15). Teachers in the intervention were provided with training; one-to-one, bi-monthly coaching; and Story Maps that included target vocabulary, open-ended questions to promote conversations during book reading, and suggested extension activities that support use of target vocabulary. The results suggested that teachers in the intervention increased on the global quality of their instruction, as well as on their fidelity to the project's strategies and their use of target vocabulary words. In addition, children in the intervention classrooms performed significantly better on measures of taught vocabulary words, and HLM analyses found gains on the Peabody Picture Vocabulary Test-4 (d = 0.19) and the Expressive One Word Picture Vocabulary Test-4 (d = 0.14), both standardized measures of vocabulary development. The results suggest that Story Talk holds promise as a relatively resource-conservative PD intervention that can be implemented with fidelity and can significantly improve children's vocabulary development, especially among children in high-poverty schools.

Research has clearly documented the critical role that vocabulary development plays in children's success in learning to read and as a result, their success in school (Dickinson, Golinkoff, & Hirsh-Pasek, 2010; Muter, Hulme, Snowling, & Stevenson, 2004). There is also considerable evidence that there are statistically significant and conceptually meaningful differences between children in poverty and their more affluent peers in the number and complexity of words children know by the time they enter preschool (Fernald, Marchman, & Weisleder, 2013; Hart & Risley, 1995). This socio-economic vocabulary disparity, spanning at least one standard deviation on commonly used vocabulary measures (Lee & Burkam, 2002), has been identified as one of the significant factors contributing to the enduring achievement gap that separates children in poverty from their middle- or upper-income peers throughout the K-12 years and beyond (Hollich, Hirsh-Pasek, & Golinkoff, 2000).

In an effort to ameliorate this disparity, much attention has been focused on understanding how children learn words and what instructional strategies can be implemented in early childhood classrooms to effectively build vocabulary skills among young children in poverty (Hindman, Wasik, & Snell, 2016). Mounting evidence from the professional development literature suggests that individualized coaching can be an effective support for teachers of young children, as it can increase the amount and quality of classroom vocabulary instruction and elevate children's skills (Hindman & Wasik, 2011). However, coaching can be resource intensive. This study describes Story Talk, a streamlined, scalable professional development model with relatively limited coaching that guides preschool teachers to employ books and classroom activities to teach vocabulary to young children.

1. How young children develop vocabulary

Children begin learning language at birth and, by the age of 24 months, have an expressive vocabulary of 200–300 words on average (Pence & Justice, 2008). Broadly speaking, children build vocabulary through meaningful, reciprocal conversations with more expert users of their language(s), particularly when rare or sophisticated vocabulary words are used (Dickinson & Tabors,

☆ This project was funded through an Institute of Education Sciences grant to the first and second authors. Grant # R305A140034.

* Corresponding author.
E-mail address: bwasik@temple.edu (B.A. Wasik).

https://doi.org/10.1016/j.ecresq.2018.11.001
0885-2006/

2001; Hoff, 2014). However, not all conversations are created equal, and the empirical literature on the mechanisms of word learning isolates four conditions that make young children more likely to learn and remember new words.

First, with the relatively rare exception of fast mapping (Carey & Bartlett, 1978), children need repeated exposures to new words in order to learn them (Biemeller & Boote, 2006; Hoff, 2003). Studies with adults suggest a range of 40 to as many as 200 exposures may be needed to learn a word (e.g., Clay, Gill, Glynn, McNaughton, & Salmon, 2007). However, simply hearing new words over and over is not likely to be sufficient. Instead, a second condition that supports word learning is the presence of an explicit definition of novel words (Biemiller, 2001; Biemiller & Boote, 2006; Weizman & Snow, 2001), especially for children with more limited language skills (Penno, Wilkinson, & Moore, 2002). These definitions might include explanations of the function of an object (Booth, 2009), as well as pictures and props (Wasik & Bond, 2001; Han, Moore, Vukelich, & Buell, 2010).

Third, children also learn more new words when they are presented in a meaningful context or theme (Harris, Golinkoff, & Hirsch-Pasek, 2011; Neuman & Dwyer, 2009). This finding is consistent with the long-held and robust discovery in memory research that information is best learned when integrated around a story or concept, as opposed to delivered as a set of isolated facts (Beck, McKeown, & Kucan, 2013; Bransford & Johnson, 1972). As an example, Christie and Roskos (2006) found that children better recalled and used vocabulary related to building materials (e.g., hammer, hard hat, tool belt) when they were presented in the context of a theme on building, rather than when they were presented in isolation.

Fourth, children learn vocabulary best from language exchanges with adults who encourage them to talk and then provide meaningful feedback on their remarks, thereby scaffolding children's linguistic and cognitive development (Smith, Landry, & Swank, 2000; Weisleder, & Fernald, 2013). For example, Cristofaro and Tamis-LeMonda (2012) examined the language exchange between three-year-olds and their mothers, finding that a higher frequency of "wh" questions was linked to increases in the amount of child talk and, ultimately, in children's vocabulary knowledge. Rowe, Leech, and Cabrera (2017) found similar results when examining the conversations between two-year-olds and their fathers. Parallel work has uncovered consistent patterns in classrooms; for example, Justice, Jiang, and Strasser (2018) found teachers' communication-facilitating behaviors that engender child conversation participation were the best predictor of children's vocabulary development. Other research supports this trend (Ard & Beverly, 2004; Hindman, Wasik, & Erhart, 2012).

A relatively smaller but emerging body of work has explored the import of the type(s) of feedback that adults provide to children's responses to questions. For example, Dickinson and Porche (2011) determined that when adults' feedback during book reading corrected or clarified children's responses, children's vocabulary grew. In addition, recasting children's responses by either repeating or expanding on what children have said also builds children's vocabulary (Baker & Nelson, 1984; Farrar, 1992). Interestingly, Cabell, Justice, McGinty, DeCoster, and Forston (2015) determined that a small number of conversations rich with interaction (i.e., adult questions, child responses, and adult feedback) was more predictive of vocabulary learning than multiple conversations with fewer exchanges, even though the total dosage of child talk was comparable. This evidence may suggest that children's vocabulary development is particularly aided by in-depth, back-and-forth discussion of a topic.

In sum, research has clearly demonstrated that providing children with multiple, meaningful exposures to words through ongoing, reciprocal interactions supports children's vocabulary learning. As a result, there has been much attention to training teachers to implement word learning strategies in preschools serving children in poverty in order to close the word gap.

2. Effective vocabulary interventions in high-need preschools

Although much is known both theoretically and empirically about how children learn words (Gleitman & Fisher, 2007; Harris et al., 2011; Pinker, 1999), remarkably less is clear regarding how to effectively translate these principles into feasible, replicable classroom instructional practices in early school and care settings. In fact, preschool and other early education contexts include many constraints that limit opportunities for extended teacher–child interactions (Gest, Holland-Coviello, Welsh, Eicher-Catt, & Gill, 2010), including high child–teacher ratios (NAEYC, 2017) and curricula that rarely prioritize vocabulary and language exchanges (Neuman & Dwyer, 2009). Consequently, many well-designed and implemented interventions have engendered gains in teacher practice but not in children's vocabulary, as evidenced by reports from the Early Reading First program (Jackson et al., 2011; Markussen-Brown et al., 2017), the Preschool Curriculum Evaluation Research program (PCERC, 2008), and the Head Start Impact Study (U.S. Department of Health and Human Services, 2010). These null effects likely emerged because these educator-support models did not produce thorough and enduring changes in teachers' application of these four core principles of word learning to their teaching (Hindman, Snell, & Wasik, 2017).

Of particular value for the field are program effects on both taught or target words (i.e., the words that were specifically focused on in the intervention) and standardized vocabulary (i.e., the broader universe of words tapped by norm-referenced tools). Gains in target words suggest that children are learning the words highlighted in the classroom, which is critically important. However, gains in standardized vocabulary imply that increased knowledge of taught words might ultimately equip children with powerful new background knowledge, which in turn that helps them attend to and make sense of additional novel words that they encounter. Indeed, for children in poverty, gains on standardized vocabulary scores reflect a narrowing of the socioeconomic word gap, because children are growing more rapidly than maturation alone would predict. However, to date, only three comprehensive PD approaches have demonstrated effects on children's taught word as well as standardized vocabulary. All can be characterized as multi-pronged PD models targeting a comprehensive set of teacher practices and child outcomes.

2.1. Comprehensive PD models

Several interventions have focused on multiple components of language and literacy, including vocabulary development, phonemic awareness, and letter learning, often using book reading as a vehicle to teach these skills and including other content areas such as math and science. First, the Texas School Ready model (see Landry, Swank, Smith, Assel, & Gunnewig, 2006; Landry, Anthony, Swank, & Monseque-Bailey, 2009) for details on the Texas School Ready program) provides preschool teachers with online training in content that includes oral language, vocabulary development, phonemic awareness, read-alouds, writing, and language and literacy progress monitoring. In addition, Texas School Ready teachers receive individualized coaching each week. In a randomized controlled trial (RCT), preschool children whose teachers received Texas School Ready training over one or two years significantly increased their performance on the Expressive Vocabulary Test compared to children in business-as-usual classrooms, ($d = 0.19$).

As another example, ExCELL (Exceptional Coaching for Early Language and Literacy) focused on similar vocabulary and literacy content for teacher professional development and used book reading as a critical component in teaching these skills. Teachers were provided with monthly workshops and weekly coaching from an expert mentor as well as access to a progress monitoring tool. In an RCT, Head Start children in ExCELL classrooms performed significantly better on standardized receptive vocabulary measures compared to children in comparison classrooms (d = 0.41) (Wasik & Hindman, 2011). And broader still, the Boston project (Weiland & Yoshikawa, 2013, 2014) implemented the Opening the World of Learning curriculum (OWL; Schickedanz & Dickinson, 2005) – which covers mathematics, literacy (with a focus on book reading), social-emotional development, science, and the arts – along with one-to-one coaching and ongoing assessment in Boston preschools. Here, too, using a regression discontinuity model, a significant effect for children whose teachers received the PD was found on the Peabody Picture Vocabulary Test-III (d = +.45) (Weiland & Yoshikawa, 2013).

While these models are effective, it is important to note that the comprehensive nature of these interventions makes them challenging to implement. For example, all three models (Texas School Ready, ExCELL, and the Boston project) included face-to-face individualized coaching (ranging from weekly to monthly) for teachers and monthly training, which becomes quite costly as it requires identifying, training, and employing high-quality coaches over time. These interventions also focused on many content areas, often requiring teachers to re-structure almost all aspects of their daily schedules. This approach demands a substantial investment of time and energy and represents a potential source of logistical challenge and stress. Particularly because these programs are most needed in the contexts with fewest resources, it is important to consider how effective interventions could be streamlined in ways that help teachers build children's vocabulary using techniques that are maximally resource-efficient. Perhaps this explains the finding from Markussen-Brown et al. (2017)'s meta-analysis that PD interventions, on average, do not result in large or significant outcomes for child language.

2.2. PD models targeting book reading

Other PD models have taken a more targeted approach to building just children's vocabulary, often supporting teachers' shared book reading as the key instructional mechanism, with intended extension into other, related classroom areas such as center activities and circle time. Books are, in theory, uniquely linked to vocabulary, as they provide opportunities for children to experience vocabulary words that they may not have the opportunity to hear and use in everyday conversations (Snow & Tabors, 1993), with images and a supporting text that help to elucidate the meaning(s) of these new words (Ganea, Picard, & DeLoache, 2008). For example, McKeown and Beck (2014) trained teachers to ask open-ended questions and explicitly define the target vocabulary words while they re-read books, and then to engage children in book-related follow-up activities to help reinforce word learning. The authors reported that children in the intervention group learned significantly more of the taught words than children in the business-as-usual control group. Similarly, Coyne, McCoach, Loftus, Zipoli, and Kapp (2009) and Simmons et al. (2011, 2008) conducted a series of studies in which researchers read books to children in small groups. During book reading, the researchers explicitly defined target words, asked children questions about the words and book, re-read each book three times, and engaged children activities that promoted use of the vocabulary words. Children in the intervention group performed significantly better than children in the comparison condition on the number of target words they knew (d = 0.8).

However, on the whole, these more streamlined, book-focused models have yielded few effects on standardized word learning (Marulis & Neuman, 2010; Wasik, Hindman, & Snell, 2016). Given the extraordinary variation in the methods of these studies (Wasik et al., 2016), it is difficult to ascertain precisely what strategies are absent from these more discrete interventions. However, generally, few of these models employed any ongoing (i.e., multiple months) individualized educator coaching, and in fact many were implemented entirely by researchers. The increase in children's word-learning opportunities throughout the classroom, then, may not have been sufficient to produce these broader gains. It is thus possible that a highly focused, intensive intervention aiming to help preschool teachers build vocabulary through book reading specifically could employ coaching, albeit to a lesser degree than is needed in more comprehensive interventions, while still yielding effects on teachers' practices and children's taught and standardized vocabulary.

3. Aims of the current study

As stated above, a relatively small number of professional development projects have successfully raised both the quality of teachers' instruction and the extent of children's vocabulary knowledge during the preschool period, particularly when child outcomes are assessed using standardized measures. One aim of the current study was to determine whether a novel PD intervention, entitled Story Talk, that offloads some of the support generally provided through coaching onto more resource-efficient materials might demonstrate teacher and child benefits. We intentionally explored effects of Story Talk on several facets of teachers' practices that are commonly examined in the literature, including, most broadly, global instructional quality, as well as fidelity to the specific practices recommended by the intervention, and, most fine-grained, the precise nature of classroom talk about vocabulary words. We also intentionally captured several constructs related to children's vocabulary, including taught words and standardized (both expressive and receptive) vocabulary scores.

A second aim was to examine the mechanics of Story Talk's effects by identifying precisely which teacher practices affected by the intervention explained gains in children's skills. To date, effective interventions generally target a variety of teaching practices, obscuring precisely what teacher instructional behaviors lead to gains in children's outcomes. As we begin to narrow down our understanding of the particular tools and experiences that constitute effective PD, we also need to refine our awareness of the highest leverage teacher practices for vocabulary development. Findings related to this latter point would further inform the aims and methods of teacher PD. The empirical literature reviewed above suggests that the key practices are likely to include teacher–child conversations (including teachers' use of open-ended questions and their feedback to children's responses), and repeated teacher and child use of target vocabulary (Justice et al., 2018; Zucker, Cabell, Justice, Pentimonti, & Kaderavek, 2013). Therefore, the current study expands beyond simply evaluating the effects of the intervention on teachers and children to also examine the unique links between each of these teacher practices and children's vocabulary learning, in an effort to map the pathways of change.

3.1. Current study research questions

Given our aims, this study explores three clusters of research questions:

Question 1: Does Story Talk raise the quality of teachers' instruction? Outcomes of interest include fidelity to the intervention procedures, global quality of teachers' practices, and the frequency of vocabulary-building teacher and child talk in the classroom.

Question 2: Does exposure to Story Talk improve children's vocabulary? Outcomes of interest include taught words from the curriculum as well as receptive vocabulary and expressive vocabulary on standardized measures.

Question 3: Which specific Story Talk instructional strategies best explain children's vocabulary outcomes? We explore global quality, fidelity to the intervention practices, and classroom talk, relating all three to (standardized) receptive and expressive vocabulary.

4. Method

4.1. Procedure

4.1.1. Recruitment

Teachers were recruited to participate in this project through an open email invitation to all preschool teachers sent from the district's central administration. Interested teachers received a visit from a member of the project staff who explained the PD in detail and explained that teachers would be assigned, at the level of their school, to either the intervention or the business-as-usual comparison, and that this assignment would be completely random. A total of 35 classrooms within 15 schools agreed to participate.

4.1.2. Design

Random assignment was conducted at the level of the school, so that all classrooms in a school shared the same assignment. This approach lessened the possibility of contamination across conditions, particularly as teachers in this district planned instruction together in grade-wide teams. Twenty classrooms were randomly assigned to the intervention and 15 were assigned to the control.

4.1.3. Intervention

The Story Talk intervention provides whole-class instruction in vocabulary, drawing on book- and play-based activities. Story Talk classrooms are provided with materials, training, and progress monitoring of children, all focused on a series of target words.

4.1.3.1. Materials. Story Talk is organized around 10 common preschool themes, each designed to cover 3 – 4 weeks of daily instruction. Story Talk also includes approximately 10 high-quality trade books related to each theme, and each book is designed to be read three times over non-consecutive days. For each book, there is a Story Map that contains target words, questions, and center activities.

A Story Map, which supports teachers' reading of the book, was developed for each book. Each Story Map begins with a *Vocabulary Introduction*, with a list of the 5 target vocabulary words and child-friendly definitions and picture cards (with an image representing the word) for each target word. Two criteria were employed in selecting target words: words had to be (a) potentially unfamiliar to the children and (b) essential for comprehension of the story. Words were a combination of tier 1, 2 and 3 words (Beck et al., 2013), with about 80% of words from tiers 1–2 and about 20% from tier 3.

Story Maps also include *Story Starters*, or 3–4 open prompts that promote pre-reading discussions and build background knowledge, particularly around the key words (e.g., Whitehurst et al., 1994). Next, Story Maps provide approx. 3–5 "During Reading" and "After Reading" open prompts targeting the key words and creating opportunities for teacher–child dialogue. Finally; *Story Extensions* suggest classroom center extension activities for five centers (Dramatic Play; Writing; Science; Math/manipulatives and Blocks) and offer questions and comments for teachers to use in those center activities to invite children to talk about the story vocabulary and concepts. These center activities offer teachers the chance to work with small groups of children and to individualize instruction (although, as above, Story Talk is fundamentally a Tier 1 intervention). Story Maps include three different versions of this content for each book to ensure that 3 *Repeated Readings* provide systematically scaffolded interactions with the book's vocabulary and ideas. In sum, Story Maps are not scripts; instead, they model research-based practices for teachers to use with all students with room for adjustment and adaptation. Consequently, while we aimed to increase the number of repetitions of target words that all children received, classrooms were likely to vary in how and how often teachers talked about the words and invited children to talk about the words (see results of Research Question 1).

4.1.3.2. Training. Teachers receive four, 3-hour sessions of group training distributed across the school year. Trainings target how to (a) implement the Story Maps (including book reading and center activities) with fidelity in their classrooms, (b) interpret and use progress monitoring data, (c) effectively encourage conversations with children in order to extend their language and vocabulary, and (d) effectively manage classroom discourse, including through turn taking and active listening (Melis, Grocke, Kalbitz, & Tomasello, 2016). In addition, individualized coaching of teachers is conducted twice per month (every other week), including an onsite observation for approximately 60 min of the instructional day (generally in the morning) during teachers' book reading and center extension activities. The coach videotapes the instruction and uses the teacher fidelity measure (described below) and field notes to gauge what teachers are doing well and to identify problem areas. Ideally in the afternoon of the same day as the visit, the coach conferences with teachers for about 30 min to provide feedback on teachers' performance on the fidelity tool, including the extent to which they demonstrated each target behavior, and also provides support for improving implementation.

4.1.3.3. Progress monitoring. A project-aligned child progress monitoring (PM) tool measures children's ongoing receptive and expressive knowledge of target vocabulary from the Story Maps (see Section 4.3). The PM measure is implemented three times per year and the data are discussed during coaching to foster teachers' data-informed instructional decision making. Teachers then use the data to determine what content to focus on with the whole class and with particular children, as well as how to group children for small-group activities.

4.1.4. Control

Teachers in the control condition received the same books as the intervention teachers but not the Story Maps. Teachers in the control condition attended district PD instead of group trainings for the same amount of time. Teachers in the control condition did not receive the coaching.

4.1.5. Data collection

Teachers in both conditions were videotaped for a morning of instruction by a coach (about 120 min) in fall of the school year, before any PD began, and again in the spring after all PD was completed. These videos were coded for analysis with a project-specific fidelity measure and the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008). The trained assessor who scored the videos was blind to condition. Children were assessed on several standardized vocabulary measures by assessors who were blind to condition. Three times during the school year (Nov.,

Feb., and Apr.), children's knowledge of taught words was assessed through the Progress Monitoring assessment.

4.1.6. Incentives

At the end (i.e., spring) of the year, all teachers (intervention and control) received $500 for their participation. Teachers also kept all classroom materials (i.e., the books, and for the intervention, iPads and Story Maps). Finally, after data collection concluded, teachers in the control condition received the Story Maps.

4.2. Participants

4.2.1. Teachers

Thirty-five public preschool teachers in an urban, high-poverty school district in the Northeast participated in this study. Based on teachers' self-reports, 42% were African American, 3% were Asian, 36% were white, 3% were Hispanic/Latino, and 16% were of other backgrounds. In addition, 100% held a Bachelor's degree, and 50% had Master's degrees. On average, teachers had 12 years of teaching experience (SD = 8.31, range = 1–32). Thirty-three of the teachers were female. Ultimately, 20 teachers were randomly assigned to the Story Talk intervention condition and 15 were assigned to the control condition.

4.2.2. Children

All children in each teacher's classroom were invited to participate in the study. Ultimately, a total of 519 children returned consent forms signed by their families, approximately 82% of the entire potential sample. The age range among children in the fall of the study was 44–76 months, with an average age of 55 months. The sample was evenly divided by gender. Six percent of children were dual language learners (predominately speaking Spanish at home). Community demographics show that the majority (83%) of children are African American, while 8% are white and 9% are Hispanic/Latino. In addition, 87% of children receive free or reduced lunch, and 15% of children have identified special needs. After schools were randomly assigned to conditions, 312 of these children were assigned to Story Talk.

4.3. Measures

4.3.1. Instructional fidelity to Story Talk

Teachers' fidelity to intended Story Talk techniques was gauged using a project-aligned measure. Items addressed teachers' adherence to each part of the Story Maps: Vocabulary Introduction, Story Starters, During-reading Questions, After-reading Questions, and Story Extensions in centers. The tool was used to code the fall and spring videotapes of teachers' classroom instruction. All segments of the morning instruction period (e.g., work at a math center, small-group activities) were potentially relevant for children's vocabulary learning and were coded for the current study.

4.3.1.1. Vocabulary Introduction. For the Vocabulary Introduction, teachers were given up to 3 points for each of 5 target vocabulary words: 1 point for introducing the word, 1 point for providing a child-friendly definition of the word, and 1 point for showing a picture of the vocabulary word. Teachers could receive a maximum of 15 points for this section.

4.3.1.2. Story Starters and During and After Questions. For the before-, during-, and after-reading questions, trained observers counted each open-ended question that teachers posed, considering the complexity of the interaction around each one. Each was considered to be worth a maximum of 8 points: 1 point for asking the question, 1 point for taking responses from 2 or more children, and up to 6 points for feedback (including 0 = no feedback, 1 = repeats child response, 2 = repeats child response with elaboration, 4 = closed follow-up question, and 6 = open-ended follow-up question).

4.3.1.3. Story Extensions. Teachers were scored on the centers they visited the day of the observation. Each center was worth 12 points: 1 point for visiting the center, 1 point for introducing the center activity, 1 point for providing the necessary materials, 1 point for children using at least 2 of the target vocabulary words in the center, 1 point for the teacher using 4 of the target vocabulary words in the center, 1 point for asking 3 open-ended questions, and up to 6 points for feedback on child responses to those questions, scored using the same feedback rubric presented for book-related questions. There was no maximum for this section, as teachers varied in the number of book-aligned centers they prepared and/or visited; teachers who prepared and visited more centers received more points, as they provided more opportunities for children to hear and use the target words.

4.3.1.4. Control teachers. Although control teachers did not have Story Maps, they were scored simply on the number of words they introduced, the number and quality of the open-ended questions they asked, and the number of book-related centers they created and visited. Thus, in theory, control teachers could amass as many points as intervention teachers.

4.3.1.5. Reliability and validity. Extensive piloting of this fidelity measure was conducted in the first two years of this project, preceding this randomized controlled trial. The study team determined interrater reliability of at least 90% agreement on item scores (through repeated rounds of having coders view and code the same video or in situ observations and then comparing scores). We also determined early in the project that intervention teachers' scores on the fidelity measure were correlated with their scores on the CLASS Instructional Support domain ($r > 0.50$).

4.3.2. Global instructional quality

Global quality of classroom instruction was assessed in fall and spring by trained data collectors using the CLASS (Pianta et al., 2008), a gold-standard observation tool that captures the overall quality of the instructional environment (e.g., Instructional Support, Emotional Support, and Classroom Organization). Although coaches collected the videos of the entire instructional morning (about 120 min), CLASS was coded by an observer blind to condition. Psychometric data across more than 3,000 classrooms show that this measure has reliability above 85% between trained raters and over multiple test sessions, and strong validity with other observation tools (e.g., ECERS-R) and with later child academic achievement.

4.3.3. Classroom talk

Spring classroom videos were coded by trained observers blind to condition to gauge the frequency of several elements of teacher and child talk. The coding scheme was developed by the authors, who then trained two research assistants (RA) to code the videos. The RAs worked together and with the authors to code each video multiple times, ensuring that, for each video, all viewers agreed on each code. As with the fidelity measure, because our focus was vocabulary, we coded all parts of the instructional morning.

4.3.3.1. Open-ended prompts. First, we coded for the frequency of open-ended prompts, defined as questions or statements that invited children to offer an idea, and for which there was no single right answer. In addition, these prompts generally required children to use more than one or two words to provide their answer. For example, teachers might ask, "What do you see on the cover of

this book?" or "Tell me what you would do next if you were this character." Each instance of an open-ended prompt received one point.

4.3.3.2. Child responses. Children's responses to prompts were coded. Each instance of a child responding to a teacher's prompt and being recognized by the teacher was assigned one point; we did not score remarks in which a child yelled out, unsanctioned, and the teacher strategically ignored (or actually did not notice) the remark, because (a) these remarks did not become part of the classroom conversation and it not clear that other children noticed (or learned from) these remarks, and (b) they were frequently unintelligible. Importantly, we defined a choral response (in which all or many children offered a response at once) as one response. In this way, more frequent child responses reflected more sanctioned conversational openings for children to offer their own thoughts.

4.3.3.3. Teacher simple feedback. We coded each instance in which a teacher offered a child simple feedback on a response, including "That's great," or "Good job," or "OK." The aim of this code was to capture when the teacher acknowledged the remark but did not provide any new information or invite additional child talk. Teachers received one point for each instance of simple feedback provided.

4.3.3.4. Teacher extended feedback. We also coded each instance in which a teacher offered a child more than one or two words as a response, including any elaboration on the child's idea. For example, if a child noted, "I saw a bee just like in the book at the park," extended feedback might involve a question about the park incident or information about why bees might prefer parks as a habitat. Teachers received one point for each instance of extended feedback provided.

4.3.3.5. Teacher use of vocabulary. Finally, we coded any situations in which the teacher used the target vocabulary words for the lesson. For teachers using Story Maps, the target vocabulary was specified by the Maps. For control teachers, we collected information about target words from the teachers' own lesson plans. We counted the number of instances in which the teacher used each target word throughout the entire, approximately 120-min morning video.

4.3.3.6. Child use of vocabulary. Parallel to the count of teachers' use of vocabulary, we counted the number of times that children used the target vocabulary words. If one child used a word, we counted this as one instance; if all children used the word chorally, we also counted this as one instance. This value was aggregated to the classroom level given the difficulty of identifying each child in the classroom during the instructional period.

4.3.4. Taught words

Project-specific progress monitoring was employed three times during the year (in both intervention and control classrooms) to measure children's expressive and receptive target word knowledge. The PM measure began by asking children to (expressively) label images of each of the target words (using images different from the picture cards). PM images include objects (e.g., cow), a face with expression (e.g., worry), or pictures of someone performing simple actions (e.g., typing), similar to the Expressive One-Word Picture Vocabulary Test and the Peabody Picture Vocabulary Test. Coaches delivered the assessment in Nov., Feb., and Apr. At each time point, a random subsample of 10 children was identified in each classroom and then individually assessed by the coach. Because children were randomly selected each time, the sample of assessed children differed at each time point. We chose this strategy so as to (a) reduce the burden of assessment on individual children, (b) reduce the burden on teachers of identifying specific children at each time point for assessment, and (c) ensure that we were able to test the PM approach with a wide array of preschoolers to gauge its feasibility.

The Nov. assessment included 15 randomly chosen key words chosen from recent themes, while the Feb. and Apr. assessments included 18 randomly chosen words, as children were older. Thus, the words selected at each time point differed as well (but, at a given time point, the same words were presented to all intervention and control classrooms). Because intervention and control classrooms followed the same district curriculum, they addressed the same themes and words on the same schedule, making the progress monitoring tool an authentic and relevant task in both contexts.

The measure included two consecutive tasks. For the first, expressive task, each child first saw an image representing one of the target words (e.g., an image of an anchor). The assessor asked, "What is this?" and the child's response was scored as follows: 3 points for correct answer (e.g., anchor), 1 point for answers that were close but not the exact word (i.e., hook) and 0 points for an incorrect answer (e.g., boat). After the child viewed all of the selected target words in the expressive task, the assessor next again presented each target word for which the child had offered an incorrect (i.e., 1- or 0-point) expressive answer in a second, receptive context. Here, that target word's image was presented with three other foil images (one from the same theme as the target word, one representing a word sounding similar to the target word, and one that looked similar to the target word's image). The assessor asked, "Show me [target word]." For the receptive scoring, children received 1 point for a correct response and 0 for an incorrect response.

The assessment required fewer than 10 min per child, and total possible scores ranged from 0 to 45–54 (depending on the total number of words). Analyses employed percentage correct to account for the fact that the number of words increased across time points. In addition, children's scores were aggregated to the classroom level to account for the fact that different children were assessed at each time point.

4.3.5. Receptive vocabulary

Child receptive language was individually assessed with the Peabody Picture Vocabulary Test-4 (PPVT- 4; Dunn & Dunn, 2007), a gold-standard early childhood measure which asks children to identify one image out of four that best represents a word given by the experimenter. Internal reliability in the standardization sample ranged from 0.96 to 0.97, while test-retest reliability ranged from 0.92 to 0.96, and alternate-form reliability ranged from 0.87 to 0.93. This individually administered measure requires about 10–15 min per child. Growth score values were used in analyses to better capture change from fall to spring.

4.3.6. Expressive vocabulary

Child expressive vocabulary was individually measured with the Expressive One-Word Picture Vocabulary Test-4 (EOWPVT-4; Martin & Brownell, 2010). In this gold-standard tool, children verbally identify images presented to them one at a time by the experimenter. In the preschool standardization sample, internal consistency coefficients ranged from 0.94 to 0.95, and test-retest correlations exceeded 0.97. This measure requires 15–20 min per child. Because growth score values are not available for this measure, raw scores were used in analyses, with age as a covariate.

4.4. Analytic strategy

4.4.1. Analysis plan

For Question 1, we employed descriptive statistics and OLS regression. For Question 2, we employed OLS regression to understand how the intervention was linked to differences in classroom-average PM scores in fall, winter, and spring, net of the effects of other variables. In addition, we employed multilevel regressions (in HLM; Raudenbush & Bryk, 2002) to account for the fact that children's receptive and expressive vocabulary scores were nested within classrooms. For Question 3, we limited HLM models to intervention participants and explored how Story Talk-related strategies (global fidelity, intervention fidelity, and classroom talk) explained variability in these key outcomes.

4.4.2. Missing data

From fall to spring, one teacher, in the control condition, left the field of teaching, but that teacher's replacement was willing to participate in the study. Because of the complexity and challenge related to imputing teachers' outcome data (i.e., spring CLASS and fidelity outcomes), this classroom was removed from teacher-level analyses. However, child-level analyses – where the key outcome was child vocabulary – included the second teacher's spring scores so that children in this classroom could be retained in analyses. All teachers otherwise had complete data.

Two children received consent forms but were not assessed at either time point; these children were removed from the study. Nine children (2%) were not assessed in fall but were assessed in spring. Forty-five children (9%) were assessed in fall but not in spring. Chi-square and *t*-tests revealed no significant patterns in attrition. For example, missing data in either fall or spring was unrelated to intervention assignment ($p > 0.10$). Further, missing data in fall or spring was not linked to any child background factors, including gender, age, or dual language status ($p > 0.10$). Moreover, missing data in spring were not linked to fall PPVT or EOWPVT scores ($p > 0.10$). Taken together, evidence suggested that data may be missing completely at random. We consequently employed a maximum likelihood correction in multilevel models at level 1 (see Research Question 3 below) (Enders, 2010).

5. Results

5.1. ***Question 1: Does Story Talk*** *improve teachers' prac****tices?***

Our first research question explored how Story Talk uniquely explained change in teachers' instruction, as measured by three increasingly narrow constructs: global instructional quality, fidelity to the specific practices of the intervention, and talk with children about vocabulary. To address the question, we discuss each outcome in turn, first offering descriptive statistics for both conditions in fall and spring (see Table 1). We then provide an inferential test of the intervention's effectiveness using OLS regression, accounting for relevant teacher covariates (see Table 2).

5.1.1. Fidelity to the intervention procedures

Descriptive statistics showed that in fall, teachers assigned to the control condition who had not yet received any PD scored 39.80 points on the fidelity tool (SD = 24.59). Peers randomly assigned to the intervention condition scored 42.70 points (SD = 20.06). There were no differences across condition at pre-test, $t(33) = 0.384$, $p = 0.703$.

At post-test, however, intervention teachers scored an average of 118.55 (SD = 26.72), while control teachers scored just 43.07 (SD = 19.55). Thus, there was no change from fall to spring among control teachers, $t(13) = 0.673$, $p = 0.513$, while intervention teachers improved by a factor of 3, and this increase was highly statistically significant, $t(19) = 11.47$, $p < 0.001$. Moreover, there was a highly significant difference between conditions in spring, $t(33) = 9.23$, $p < 0.001$.

As a final step, OLS regression showed that, accounting for fall fidelity score as well as teacher education and experience, teachers in the intervention gained more than three quarters of a standard deviation on fidelity ($\beta = 0.76$, $p < 0.001$, d = 3.29) over the year, a significant gain ($p < 0.001$). Overall, then, results show that training in Story Talk raises teachers' fidelity of use of the intended practices.

5.1.2. Global instructional quality

In fall, both intervention and control teachers scored near the nation's average among preschool teachers on Instructional Support (M = 2.43, SD = 0.72 and M = 2.07, SD = 0.53, respectively), Emotional Support (M = 5.56, SD = 0.71 and M = 5.65, SD = 0.77, respectively) and Classroom Organization (M = 5.17, SD = 0.66 and M = 5.18, SD = 0.71, respectively). Intervention and control teachers showed no differences on CLASS at fall pre-test on Instructional Support, $t(33) = 1.62$, $p = 0.114$; Emotional Support, $t(33) = 0.344$, $p = 0.733$; or Classroom Organization $t(33) = 0.043$, $p = 0.966$.

However, in spring, differences were observed between conditions for Instructional Support, for which intervention teachers scored 4.28 (SD = 0.86) while control teachers scored 2.61 (SD = 0.90), a significant difference $t(32) = 5.47$, $p < 0.001$. In addition, intervention teachers scored 5.92 (SD = 0.70) on Classroom Organization, higher than control teachers (M = 5.45, SD = 0.83), a marginal difference, $t(32) = 2.00$, $p = 0.054$. No differences were observed for Emotional Support (M = 5.88, SD = 0.52 and M = 5.76, SD = 0.76, respectively), $t(21.53) = 0.50$, $p = 0.625$, equal variances not assumed ($p = .042$).

As a final step, OLS regression revealed that, accounting for fall CLASS score as well as education and experience, Story Talk teachers performed significantly better on Instructional Support ($\beta = 0.63$, $p < 0.001$, d = 1.51) and Classroom Organization ($\beta = 0.35$, $p = 0.009$, d = 0.77) at post-test compared to the control teachers. Consistent with *t* tests, no differences emerged for Emotional Support ($\beta = 0.16$, $p = 0.291$).

5.1.3. Classroom talk

We address each type of classroom talk in turn, first offering descriptive data thru *t*-tests and then inferential data from OLS accounting for teacher education, and experience.

5.1.3.1. Teacher open prompts. In spring, teachers in the control condition asked 21.40 (SD = 15.96) open-ended questions, compared to 61.50 (SD = 15.50) among intervention teachers. This difference was statistically significant, $t(33) = 7.48$, $p < 0.001$. OLS regression revealed that Story Talk teachers outperformed peers, $\beta = 0.80$, $p < 0.001$, $d = 2.55$.

5.1.3.2. Child responses. Children in the control condition offered 43.67 (SD = 33.31) responses, whereas children in the intervention condition provided 116.90 responses (SD = 32.75), a statistically significant difference, $t(33) = 6.50$, $p < 0.001$. OLS regression showed differences across conditions, accounting for background covariates, $\beta = 0.74$, $p < 0.001$, $d = 2.22$.

5.1.3.3. Teacher simple feedback. Teachers in the control condition provided 16.93 instances of simple feedback (SD = 10.45), in contrast to the 34.45 instances by intervention teachers (SD = 14.70), a statistically significant difference, $t(33) = 3.93$, $p < 0.001$. OLS regression showed differences across conditions, accounting for background covariates, $\beta = 0.59$, $p < 0.001$, $d = 1.34$.

5.1.3.4. Teacher extended feedback. Control teachers provided 21.73 instances of extended feedback (SD = 21.38), in contrast to

Table 1
Descriptive statistics and *t*-tests for teacher/classroom variables in the fall and spring.

	Treatment		Control		*d*	*t*-test
	M	SD	M	SD		
Fall						
Instructional fidelity	42.70	20.06	39.80	24.59	0.13	$p = 0.703$
Global instructional quality						
Instructional Support	2.43	0.72	2.07	0.53	0.56	p = 0.114
Emotional Support	5.56	0.72	5.65	0.77	−0.12	p = 0.733
Classroom organization	5.17	0.66	5.18	0.71	−0.02	p = 0.966
Spring						
Instructional fidelity	118.55	26.72	43.07	19.55	3.15	p < 0.001
Global instructional quality						
Instructional Support	4.28	0.86	2.61	0.90	1.90	p < 0.001
Emotional Support	5.88	0.52	5.76	0.76	0.19	p = 0.625
Classroom organization	5.92	0.70	5.45	0.83	0.62	p = 0.054
Classroom Talk						
Open-ended questions	61.50	15.50	21.40	15.96	2.56	p < 0.001
Child responses	116.90	32.75	43.67	33.31	2.22	p < 0.001
Simple feedback	34.45	14.70	16.93	10.45	1.34	p < 0.001
Extended feedback	67.25	21.06	21.73	21.38	2.15	p < 0.001
Teacher use of target vocabulary	124.60	31.06	29.47	34.97	2.90	p < 0.001
Child use of target vocabulary	46.25	14.34	6.13	7.78	3.34	p < 0.001

Table 2
Effects of the intervention on teacher outcomes from OLS regressions.

Outcome	Intervention beta	*p* Value	*d*
Global quality — CLASS Instructional Support	0.63	<0.001	1.51
Global quality — CLASS Emotional Support	0.16	0.291	0.33
Global quality — CLASS Classroom Organization	0.35	0.009	0.77
Intervention fidelity	0.76	<0.001	3.29
Teacher open-ended prompts	0.80	<0.001	2.55
Child responses	0.74	<0.001	2.22
Teacher simple feedback	0.59	<0.001	1.34
Teacher extended feedback	0.72	<0.001	2.15
Teacher target words	0.81	<0.001	2.90
Child target words	0.86	<0.001	3.34

the 67.25 instances by intervention teachers (SD = 21.06), a statistically significant difference, $t(33) = 6.29$, $p < 0.001$. OLS regression also showed a significant, unique contrition of condition, $\beta = 0.72$, $p < 0.001$, $d = 2.15$.

5.1.3.5. Teacher target words. Teachers in the control condition used target vocabulary words on 29.47 occasions (SD = 34.97), while intervention teachers used target words 124.60 times (SD = 31.06), a significant difference, $t(33) = 8.50$, $p < 0.001$. OLS regression showed differences across conditions, accounting for background covariates, $\beta = 0.81$, $p < 0.001$, $d = 2.90$.

5.1.3.6. Child target words. Children in the control condition used target vocabulary words 6.13 times (SD = 7.78), while intervention children used target words 46.25 times (SD = 14.34), a significant difference, $t(30.48) = 10.60$, $p < 0.001$, accounting for unequal variances. OLS regression showed differences across conditions, accounting for background covariates, $\beta = 0.86$, $p < 0.001$, $d = 3.34$.

5.1.3.7. Correlations among types of talk. It is interesting to note that correlations amongst these variables are quite high ($r > 0.70$ and $p < 0.001$ for all). In other words, classrooms with more of one of these vocabulary-building discourse practices had, on average, more of all of these practices.

Table 3
Descriptive statistics and *t*-tests for words taught.

Treatment		Control		*d*	*t*-test
M	SD	M	SD		
November					
64.10	7.75	49.07	5.15	0.74	$p < 0.001$
February					
70.65	7.47	57.53	6.43	1.86	$p < 0.001$
April					
65.65	7.73	51.87	5.28	2.03	$p < 0.001$

Note: This variable is analyzed at the classroom level.

5.2. Question 2: Does Story Talk improve child vocabulary?

5.2.1. Taught words

As above, because different children were assessed at each time point (November, February, and April), scores were aggregated to the classroom level; further, because more words were included in February and April, percentages correct were employed in analyses. In November (see Table 3), children in the intervention classroom scored 64.10% of the possible points (SD = 7.75), whereas children in control classrooms scored 49.07% (SD = 5.15). In February, children in intervention classrooms scored 70.65% (SD = 7.47), whereas control children scored 57.53% (SD = 6.43). In April, children in the intervention classrooms scored 65.65%, (SD = 7.73) while control children scored 51.87% (SD = 5.28). Differences between conditions were significant in November, $t(33) = 6.50$, $p < 0.001$, February,

Table 4
Descriptive statistics and *t*-tests for student level outcomes in the fall and spring.

	Treatment		Control		*d*	*t*-test
	M	SD	M	SD		
Fall						
Receptive vocabulary	119.73	16.88	115.79	18.16	0.23	$p = 0.01$
Expressive vocabulary	51.55	17.52	46.77	18.04	0.27	$p = 0.003$
Spring						
Receptive vocabulary	129.79	16.20	124.43	18.36	0.31	$p < 0.001$
Expressive vocabulary	62.00	15.61	56.58	17.95	0.33	$p < 0.001$

$t(33) = 5.48$, $p < 0.001$, and April, $t(33) = 5.93$, $p < 0.001$. As a final step, OLS regression analyses tested whether differences between conditions remained significant accounting for teacher education and experience, and differences remained for November ($\beta = 0.77$, $d = 2.53$), February ($\beta = 0.71$, $d = 2.10$), and April ($\beta = 0.73$, $d = 2.23$), $p < 0.001$ for all.

5.2.2. Standardized child vocabulary measures

We conducted multilevel analyses using HLM 7.01 because children were nested in classrooms, meaning that, in contrast to core assumptions required for regression, individuals within the same classroom would likely share variance in spring vocabulary not otherwise accounted for in models. Although classrooms were nested within schools, the total number of classrooms per school was small and did not support a unique level of analysis for a three-level model. The intervention variable was placed at level 2 (classroom), as well as teacher-level variables including master's degree, years of experience, and minority ethnicity. Child-level variables were placed at level 1, including child age, gender, and dual language status, as well as child fall score on the relevant outcome (see Table 4 for descriptives). All variables were centered at the grand mean. Standardized associations between each variable and the outcome use the ratio of the unstandardized effect from the multilevel model to the pooled standard deviation of the outcome in fall (given that the implementation of the intervention throughout the school year would likely inflate the standard deviation in spring), resulting in a metric much like Cohen's *d*.

5.2.2.1. Receptive vocabulary. The fully unconditional model showed that 8% of children's receptive vocabulary scores were nested within classrooms, $p < 0.001$. As in Table 5, accounting for child gender, dual language status, age, and fall PPVT scores, results showed that children in Story Talk made significantly greater gains on the Peabody Picture Vocabulary Test-4 ($b = 3.42$, $p = 0.007$, $d = 0.19$) than peers in the control classrooms. No significant variance remained to be explained in the final model ($p = 0.068$).

5.2.2.2. Expressive vocabulary. The fully unconditional model showed that 5% of children's expressive vocabulary scores were nested within classrooms, $p = 0.004$. As in Table 6, accounting for child gender, dual language status, age, and fall EOWPVT scores, results showed that children in Story Talk made greater gains on the EOWPVT ($b = 2.56$, $p = 0.010$, $d = 0.14$) relative to peers in the control classrooms. No significant variance remained to be explained in the final model ($p = 0.297$).

5.3. Question 3: What specific intervention-related strategies best explain child vocabulary outcomes?

Finally, we explored the extent to which each of three relevant, proximal variables related to the intervention – global quality, fidelity to the Story Talk intervention, and teachers' use of target vocabulary – might explain gains in child skills. While we initially aimed to include multiple classroom talk practices, as above, we

Table 5
Effects of intervention on receptive vocabulary.

	Coefficient	Standard error	*t*-ratio	*df*	*p*-Value
Intercept	127.70	0.55	231.61	29	<0.001
Intervention	3.42	1.18	2.89	29	0.007
Teacher is of minority ethnicity	2.39	0.95	2.51	29	0.018
Teacher has master's degree	0.34	1.16	0.29	29	0.773
Teacher years of experience	0.12	0.06	1.90	29	0.067
Child is female	0.84	0.88	0.95	409	0.344
Child is dual language learner	−0.60	1.71	−0.35	409	0.727
Child fall age	0.045	0.15	0.30	409	0.764
Child fall PPVT score	0.75	0.03	25.95	409	<0.001
Classroom variance	3.56	1.89			
Residual	101.70	10.08			

Note: χ^2 (29) = 41.05, $p = 0.068$.

Table 6
Effects of intervention on expressive vocabulary.

	Coefficient	Standard error	*t*-ratio	*df*	*p*-Value
Intercept	59.73	0.41	143.64	29	<0.001
Intervention	2.56	0.94	2.74	29	0.010
Teacher is of minority ethnicity	1.28	0.93	1.38	29	0.179
Teacher has master's degree	0.34	0.94	0.36	29	0.722
Teacher years of experience	0.13	0.048	2.75	29	0.010
Child is female	0.49	0.63	0.79	408	0.430
Child is dual language learner	−0.99	1.08	−0.91	408	0.361
Child age	0.10	0.14	0.70	408	0.486
Child fall score on EOWPVT	0.77	0.03	26.69	408	<0.001
Classroom variance	0.88	0.94			
Residual	77.95	8.83			

Note: χ^2 (29) = 32.53, $p = 0.297$.

Table 7
Dominance analysis for receptive vocabulary.

	ISD	Fidelity	Vocabulary
Receptive vocabulary			
Alone in model	0.08**		
		0.08*	
			0.10**
Pairs			
	0.04	0.06	–
	0.03	–	0.08
	–	0.04	0.06
All together	0.01	0.04	0.06
Expressive vocabulary			
Alone in model	0.06*		
		0.07*	
			0.06*
Pairs	0.03	0.05∼	
		0.05∼	0.02
	0.04		0.03
All together	0.02	0.05	0.00

*.05; **.01.

determined that they were all highly correlated and thus could not be used concomitantly; we selected teachers' use of target vocabulary because we found this practice to be most conceptually distinct from the items of the fidelity measure, which focused heavily on questions and feedback, and would therefore provide important additional information about the classroom context beyond the other measures. To gauge the relative import of these multiple predictors, we employed dominance analysis (Azen & Budescu, 2003). Accordingly, we ran a series of models, testing each of these predictors independently, in all possible pairs, and as a set of three, exploring the proportion of the variance in the outcome explained by each variable in each model. We employed as a base the model from Question 2, but we removed the intervention variable. Thus, we systematically explored the degree to which each of the potential teacher-level outcomes of the intervention was uniquely predictive of change in children's skills. It is critical to note that, among the three predictors tested in these analyses, Pearson zero-order correlations were high ($r > 0.70$ for all), meaning that the practical differences between these constructs are nuanced.

5.3.1. Receptive vocabulary

Results (see Table 7) suggest that teachers' use of vocabulary was the strongest proximal predictor of children's receptive vocabulary learning ($\beta = 0.10$, $p < 0.01$). Specifically, teachers' frequency of using target vocabulary demonstrated the strongest link to receptive vocabulary when each predictor was added to the model independently, as well as in all pairs and when all three variables were added at the same time.

5.3.2. Expressive vocabulary

Results (see Table 7) indicated that expressive vocabulary was most closely linked to the frequency and quality of teachers' use of Story Talk strategies ($\beta = 0.07$, $p < 0.05$). Specifically, teachers' fidelity scores demonstrated the strongest link to expressive vocabulary when each predictor was added to the model independently, as well as in all pairs and when all three variables were added at the same time.

6. Discussion

This study explored the degree to which Story Talk, a preschool teacher professional development (PD) intervention focused on supporting vocabulary instruction, changed the frequency and quality of teachers' practices and raised children's vocabulary skills. Findings suggest that teachers trained in Story Talk made important and vocabulary-enhancing changes in their behaviors. Specifically, teachers' instructional quality and organization increased in meaningful and significant ways over the course of the intervention. Further, looking more closely at specific instructional practices, analyses revealed that teachers in Story Talk significantly and meaningfully increased the quality and frequency of using Story Talk strategies, including highlighting vocabulary throughout their instruction, asking open-ended questions, and providing meaningful feedback to children on their language, relative to peers in the control condition. In addition, children in this high-need preschool context learned substantially more words from the intervention and began to close the achievement gap on standardized measures of receptive and expressive vocabulary. Finally, evidence potentially traced standardized receptive vocabulary gains to teachers' use of new vocabulary with children, whereas expressive vocabulary gains were better explained by a broader array of practices including asking open-ended questions and providing feedback to child language around new vocabulary.

6.1. Child vocabulary development

Most important for this study are the findings that the Story Talk intervention engendered significant increases in children's word learning, both on project-specific assessments and standardized measures. Children in intervention classrooms were assessed three times per year (Nov., Feb., and Apr.) on target vocabulary words that were presented on the Story Maps and that children heard during the book readings and from the questions on the Story Maps during book reading and center activities. Critically, children in the control condition were exposed to the same words, as they experienced the same district-wide curriculum and had the same books in their classroom. Children in intervention classrooms demonstrated stronger target word knowledge at all time points; the size of the effect was large. Moreover, the benefit of intervention participation remained relatively steady across time points, in contrast to some previous studies of preschool teacher PD (Hindman & Wasik, 2011).

Further, children in the intervention group scored significantly higher on both the Peabody Picture Vocabulary Test-4, a standardized measure of receptive vocabulary, and the Expressive One-word Picture Vocabulary Test-4, a standardized measure of expressive vocabulary. Although the words presented on these tools were not taught in Story Talk classrooms, Story Talk teachers were trained to increase the focus on vocabulary words, specifically inviting children to hear and use these new words in the context of book reading and center activities. This heightened attention to words and their meanings could have helped children become more interested in and motivated around learning new words in general, and as a result, accelerated their performance on these measures. Another possible explanation is that the increased and systematic exposure to books resulted in children encountering and learning non-target words that actually were on the standardized measures. Regardless, the positive implication of this finding is that, although children's movement on standardized vocabulary has been difficult to show (Marulis & Neuman, 2010), Story Talk children are learning more words than maturation alone would predict, closing the gap that separates these children in poverty from the skill levels of their more affluent peers. This finding, then, implies that Story Talk may be a strong classroom support for children in poverty.

Although the field cannot point to a particular target effect size on the PPVT-4 that is clinically meaningful, Cohen defines an effect of 0.2 or below as a small effect size. We suggest several possible reasons for this small (but significant) effect. One possibility is that it is difficult to increase standard scores for preschoolers because they are already significantly behind their more advantaged peers. Another possibility is that the amount (dosage) of the oral language opportunities children were engaged in, which has been identi-

fied as a significant predictor of children's language development (Justice et al., 2018), was not enough to result in a medium or high effect size. The teachers in the RCT were in their first year of training of using the Story Maps, open-ended prompts, and providing meaningful feedback. Perhaps with more time to develop expertise in delivering Story Talk, the child outcomes would improve and as a result, narrow the vocabulary gap between children in poverty and their more advantaged peers. This hypothesis needs to be tested empirically.

6.2. Teacher practices

A second critical finding from this work is that a relatively streamlined teacher professional development intervention, in terms of hours of coaching, had strong effects on teachers' practices, as measured in several, increasingly fine-grained ways.

6.2.1. Global instructional quality

Most broadly, teacher behaviors were measured by the CLASS, which includes the three domains of instructional support, classroom organization, and emotional support. Regarding Instructional Support, at pre-test, there were no differences between intervention and control teachers on their CLASS scores. However, post-test CLASS scores revealed that intervention teachers scored significantly higher on Instructional Support. Given Story Talk's emphasis on training and one-to-one coaching focused around engaging children in conversations around salient vocabulary, as well as the Story Maps that explicitly provided questions that promoted children's language, gains on this assessment tool are closely aligned with Story Talk's inputs to teachers. Moreover, significant findings emerged for Classroom Organization, which may be the result of Story Talk introducing the systematic reading of books each day, each of which was read three times (increasing children's familiarity with the content) and was followed by structured book-related center activities. In other words, the high level of structure and routine in the Maps may have engendered change on Classroom Organization, at least during this literacy-focused part of the day.

6.2.2. Fidelity to intervention processes

In addition to global quality, we carefully tracked teachers' fidelity to the intervention. Our fidelity tool differed somewhat from other tools in the field, because there was no single set of practices that reflected high fidelity. Instead, because teachers were implementing Story Maps, which were not a script, teachers needed to mediate the Story Maps for the young children in their classrooms. Consequently, teachers were scored on the number of intervention-related practices they used as well as the quality with which they used these practices. Critically, the intervention raised the frequency and quality of intervention strategy use. In turn, the use of the Story Map strategies – including introducing vocabulary, asking questions throughout book reading, and implementing book-related center activities, all through interactive conversations with children – was uniquely predictive of children's expressive vocabulary learning. This finding implies that these give-and-take conversations are critical for children's emerging skills at using new words. Moreover, this finding helps us understand what features of the intervention package might actually matter most for children's learning, often something of a "black box" with regard to multi-pronged teacher professional development models.

6.2.3. Classroom talk

Examining classroom talk, the data suggest that in the spring, intervention teachers were asking more than three times as many open-ended prompts compared to teachers in the control condition. In addition, teachers in the intervention provided more feedback, including extensions, to children's comments than teachers in the control group. Intervention teachers also used more of the target words in their interactions with children than control teachers. These are all strategies that intervention teachers were trained on and that are identified as research-based practices to support vocabulary development (Justice et al., 2018). The classroom talk findings also indicate that children in the intervention group used more target words than children in the control group. In sum, teachers in the intervention classrooms provided a richer, more vocabulary-focused language environment for children.

In examining the specific strategies teachers used that predicted children's receptive and expressive vocabulary learning, the data suggest that, for receptive vocabulary learning, teachers' explicit use of vocabulary words during book reading and center activities is the best predictor. Therefore, these findings suggest that, in particular, teachers' repeated use of target vocabulary words in meaningful contexts can promote children's broader receptive vocabulary learning and that, in Story Talk, this feature of teacher change may be particularly important

6.3. Limitations

Although this study provides important findings, limitations highlight helpful next steps for this line of work. One of the limitations of this study is the small sample of teachers. Although we did employ a randomized controlled trial to evaluate the intervention, this intervention was conducted on a small scale, and it would be useful to replicate this intervention with a larger sample size. A second limitation is that coaches videotaped classrooms for observational data and assessed children for the PM; they were not blind to condition. Future work will try to secure resources that allow for all data to be collected by personnel blind to condition. Third, we focused primarily on monolingual learners in this study. Our plan is to extend this intervention, examining the impact this approach could have with dual language learners (DLLs) and exploring the modifications that would need to be made to support these children.

7. Conclusion

In sum, the data suggest that Story Talk holds promise as an effective, research-based intervention designed to increase children's vocabulary knowledge. Vocabulary learning is one of the most significant predictors of children's success in learning to read and as a result, their success in school. Story Talk, with a streamlined approach to training and effective materials, could help close the vocabulary gap in high-poverty, preschool classrooms.

References

Ard, L., & Beverly, B. (2004). Preschool word learning during joint book reading: Effect of adult questions and comments. *Communication Disorders Quarterly*, *26*(1), 17–58. http://dx.doi.org/10.1177/15257401040260010101

Azen, R., & Budescu, D. V. (2003). The dominance analysis approach for comparing predictors in multiple regression. *Psychological Methods*, *8*(2), 129–148. http://dx.doi.org/10.1037/1082989X.8.2.129

Baker, N. D., & Nelson, K. E. (1984). Recasting and related conversational techniques for triggering syntactic advances by young children. *First Language*, *5*(13), 3–22. http://dx.doi.org/10.1177/014272378400501301

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

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

Biemiller, A. (2001). Teacher vocabulary: Early, direct, and sequential. *American Educator*, *25*(1), 24–28.

Booth, P. (2009). The development of vocabulary proficiency in relation to learning style. In Benati, & G. Alessandro (Eds.), *Issues in second language proficiency* (pp. 95–115). London, U.K: Continuum.

Bransford, J., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, *6*(11), 717–726. http://dx.doi.org/10.1016/S0022-5371(72)80006-9

Cabell, S. Q., Justice, L. M., McGinty, A. S., DeCoster, J., & Forston, L. D. (2015). Teacher–child conversations in preschool classrooms: Contributions to children's vocabulary development. *Early Childhood Research Quarterly*, 80–92. http://dx.doi.org/10.1016/j.ecresq.2014.09.004

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford child language conference*, 17–29.

Christie, J., & Roskos, K. (2006). Standards, science, and the role of play in early literacy education. In D. G. Singer, R. M. Golinkoff, & K. Hirsh-Pasek (Eds.), *Play = learning: How play motivates and enhances children's cognitive and social-emotional growth*. New York: Oxford University Press.

Clay, M., Gill, M., Glynn, T., McNaughton, T., & Salmon, K. (2007). *Record of oral language: Observing changes in the acquisition of language structures* (3rd ed.). New Zealand: Heinemann.

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. http://dx.doi.org/10.1086/598840

Cristofaro, T. N., & Tamis-LeMonda, C. S. (2012). Mother-child conversations at 36 months and at pre-kindergarten: Relations to children's school readiness. *Journal of Early Childhood Literacy*, *12*(1), 68–97. http://dx.doi.org/10.1177/1468798411416879

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

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

Dickinson, D. K., Golinkoff, R. M., & Hirsh-Pasek, K. (2010). Speaking out for language: Why language is central to reading development. *Educational Researcher*, *39*(4), 305–310. http://dx.doi.org/10.3102/0013189X10370204

Dunn, M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test-4*. MN: AGS: Circle Pines.

Enders, C. K. (2010). *Applied missing data analysis*. Baltimore: Guildford.

Farrar, M. J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology*, *28*(1), 90–98. http://dx.doi.org/10.1037/0012-1649.28.1.90

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, *16*(2), 234–248. http://dx.doi.org/10.1111/desc.12019

Ganea, P., Pickard, M., & Deloache, J. (2008). Transfer between picture books and the real world by very young children. *Journal of Cognition and Development*, *9*(1), 46–66. http://dx.doi.org/10.1080/15248370701836592

Gest, S. D., Holland-Coviello, R., Welsh, J. A., Eicher-Catt, D. L., & Gill, S. (2010). Language development subcontexts in Head Start classrooms: Distinctive patterns of teacher talk during free play, mealtime, and book reading. *Early Education and Development*, *17*(2), 293–315. http://dx.doi.org/10.1207/s15566935eed1702_5

Gleitman, L., & Fisher, C. (2007). *Language acquisition*. New York, NY: Blackwell Publishers.

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*. New York: Guilford.

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

Hindman, A. H., & Wasik, B. A. (2011). Exploring Head Start teachers' language and literacy knowledge: Lessons from the ExCELL professional development intervention. *NHSA Dialog: A Research-to-Practice Journal for the Early Intervention Field*, *14*(4), 293–315.

Hindman, A. H., Wasik, B. A., & Snell, E. K. (2016). Closing the Thirty Million Word Gap: Next steps in designing research to support practice. *Child Development Perspectives*, *10*(2), 134–139.

Hindman, A. H., Snell, E. K., & Wasik, B. A. (2017). Developing vocabulary in the early grade: Research-to-practice to support standards-based curricula at home and school. In L. B. Bailey (Ed.), *Implementing a Standards-Based Curriculum in the Early Childhood Classroom*. New York, NY: Taylor and Francis/Routledge.

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

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic statue affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1268–1378. http://dx.doi.org/10.1111/1467-8624.00612

Hoff, E. (2014). Introduction to the special section: Language development in multilingual environments. *International Journal of Behavioral Development*, *38*(4), 307–308. http://dx.doi.org/10.1177/0165025414533762

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of The Society For Research in Child Development*, *65*(3), v-123.

Jackson, R., McCoy, A., Pistorino, C., Wilkinson, A., Burghardt, J., Clark, M., . . . & Swank, P. (2011). *National evaluation of early reading first: Final report (NCEE 2007–4007)*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Justice, L. M., Jiang, H., & Strasser, K. (2018). Linguistic environment of preschool classrooms: What dimensions support children's language growth? *Early Childhood Research Quarterly*, *42*, 79–92. http://dx.doi.org/10.1016/j.ecresq.2017.09.003

Landry, S. H., Swank, P. R., Smith, K. E., Assel, M. A., & Gunnewig, S. B. (2006). Enhancing early literacy skills for preschool children: Bringing a professional development model to scale. *Journal of Learning Disabilities*, *39*(4), 306–324. http://dx.doi.org/10.1177/00222194060390040501

Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology*, *101*(3), 448–465. http://dx.doi.org/10.1037/a0013842

Lee, V., & Burkam, D. (2002). *Inequality at the starting gate*. Washington, DC: Economic Policy Institute.

Markussen-Brown, J., Juhl, C. B., Piasta, S., Bleses, D., Højen, A., & Justice, L. (2017). The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly*, *38*, 97–115. http://dx.doi.org/10.1016/j.ecresq.2016.07.002

Martin, N. A., & Brownell, R. (2010). *Expressive One-Word Picture Vocabulary Test –4 (EOWPVT-4)*. Novato, CA: Academic Therapy Publications.

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. http://dx.doi.org/10.3102/0034654310377087

McKeown, M. G., & Beck, I. L. (2014). Effects of vocabulary instruction on measures of language processing: Comparing two approaches. *Early Childhood Research Quarterly*, *29*(4), 520–530. http://dx.doi.org/10.1016/j.ecresq.2014.06.002

Melis, A. P., Grocke, P., Kalbitz, J., & Tomasello, M. (2016). One for you, one for me: Humans unique turn-taking skills. *Psychological Science*, *27*(7), 987–996. http://dx.doi.org/10.1177/0956797616644070

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, *40*(5), 665–681. http://dx.doi.org/10.1037/0012-1649.40.5.665

National Association for the Education of Young Children. (2017). *Early learning standards*.. Retrieved from: https://www.naeyc.org/sites/default/files/globallyshared/downloads/PDFs/accreditation/earlylearning/Standards%20and%20Accreditation%20Criteria%20%26%20Guidance%20for%20Assessment_April%202017_3.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/10.2307/41228653

Pence, K. L., & Justice, L. M. (2008). *Language development for theory to practice*. Upper Saddle River, NY: Pearson Education, Inc.

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

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system (CLASS) manual: Pre-K*. Baltimore, MD: Brookes.

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York, NY: Basic Books.

Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of preschool curriculum on school readiness (NCER 2008–2009)*. Washington, DC: National Center for Educational Research, Institute of Education Sciences, U.S. Department of Education.

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

Rowe, M. L., Leech, K. A., & Cabrera, N. (2017). Going beyond input quantity: Wh-questions matter for toddlers' language and cognitive development. *Cognitive Science*, *41*(1), 162–179. http://dx.doi.org/10.1111/cogs.12349

Schickedanz, J., & Dickinson, D. K. (2005). *Opening the world of learning: A comprehensive literacy program*. Parsippany, NJ: Pearson Early Learning.

Simmons, D. C., Coyne, M. D., Kwok, O., McDonagh, S., Harn, B. A., & Kame'enui, E. J. (2008). Indexing response to intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities*, *41*(2), 158–173. http://dx.doi.org/10.1177/0022219407313587

Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O., Simmons, L., Johnson, Z. M., & Crevecoeur, Y. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children*, *77*(2), 207–228. http://dx.doi.org/10.1177/001440291107700204

Smith, K. E., Landry, S. H., & Swank, P. R. (2000). The influence of early patterns of positive parenting on children's preschool outcomes. *Early Education and Development*, *11*(2), 147–169. http://dx.doi.org/10.1207/s15566935eed1102_2

Snow, C. E., & Tabors, P. O. (1993). Language skills that relate to literacy development. In B. Spodek, & O. Saracho (Eds.), *Yearbook in early childhood education* (Vol. 4) (pp. 1–20). New York: Teachers College Press.

U.S. Department of Health and Human Services Administration for Children and Families. (2010). *Head Start impact study: Final report*. Washington, DC: Government Printing Office.

Wasik, B. A., & Bond, M. A. (2001). Beyond the pages of a book: Interactive book reading in preschool classrooms. *Journal of Educational Psychology*, *93*(2), 243–250.

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology*, *103*(2), 455–469. http://dx.doi.org/10.1037/a0023067

Wasik, B. A., Hindman, A. H., & Snell, E. K. (2016). Book reading and vocabulary development: A systematic review. *Early Childhood Research Quarterly*, *37*, 39–57.

Weiland, C., & Yoshikawa, H. (2013). The impacts of an urban public prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development*, *84*(6), 2112–2130. http://dx.doi.org/10.1111/cdev.12099

Weiland, C., & Yoshikawa, H. (2014). Does peer socio-economic status predict children's gains in receptive vocabulary and executive function in prekindergarten? *Journal of Applied Developmental Psychology*, *35*(5), 422–432. http://dx.doi.org/10.1016/j.appdev.2014.07.001

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. http://dx.doi.org/10.1177/0956797613488145

Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*(2), 265–279. http://dx.doi.org/10.1037/0012-1649.37.2.265

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., et al. (1994). A picture book reading intervention in day care and home for children from low-income families. *Developmental Psychology*, *30*(5), 679–689.

Zucker, T. A., Cabell, S. Q., Justice, L. M., Pentimonti, J. M., & Kaderavek, J. N. (2013). The role of frequent, interactive prekindergarten shared reading in the longitudinal development of language and literacy skills. *Developmental Psychology*, *49*, 1425–1433. http://dx.doi.org/10.1037/a0030347