© 2019, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be published in the J*ournal of Educational Psychology* and will be available, upon publication, via its DOI: 10.1037/ edu0000380

At-Scale, State-Sponsored Language and Literacy Professional Development: Impacts on Early Childhood Classroom Practices and Children's Outcomes

Shayne B. Piasta, Kristin S. Farley, Susan A. Mauck, Pamela Soto Ramirez, Rachel E. Schachter, Ann A. O'Connell, Laura M. Justice, Caitlin F. Spear, and Melissa Weber-Mayrer

The Ohio State University

Author Note

Shayne B. Piasta, Department of Teaching and Learning and Crane Center for Early Childhood Research and Policy, The Ohio State University; Kristin S. Farley, Department of Teaching and Learning and Crane Center for Early Childhood Research and Policy, The Ohio State University; Susan A. Mauck, Department of Educational Studies, The Ohio State University; Pamela Soto Ramirez, Department of Educational Studies, The Ohio State University; Rachel E. Schachter, Crane Center for Early Childhood Research and Policy, The Ohio State University; Ann A. O'Connell, Department of Educational Studies, The Ohio State University; Laura M. Justice, Department of Educational Studies and Crane Center for Early Childhood Research and Policy, The Ohio State University; Caitlin F. Spear, Crane Center for Early Childhood Research and Policy, The Ohio State University; Melissa Weber-Mayrer, Department of Teaching and Learning and Crane Center for Early Childhood Research and Policy, The Ohio State University.

Kristin S. Farley is now at the Education Department, Wittenberg University. Rachel E. Schachter is now at the Department of Child, Youth and Family Studies, University of Nebraska. Caitlin F. Spear is now at the Office of Child Development, University of Pittsburgh. Melissa Weber-Mayrer is now at the Ohio Department of Education.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305E100030 to The Ohio State University. This research would not have been possible without collaboration from the Early Childhood Quality Network, including Dennis Sykes, Sharon Sullivan, Kathryn Rider, and Melissa Ross, or the cooperation of the Ohio Department of Education's Office of Early Learning and School Readiness. We also express deep appreciation for the early childhood agencies, administrators, educators, and PD facilitators involved, as well as the research staff members responsible for coordinating and executing data collection and other research activities. The opinions expressed are those of the authors and do not necessarily represent the views of the Institute of Education Sciences, U.S. Department of Education, Early Childhood Quality Network, or Ohio Department of Education. Results were presented at the Japanese-American-German Frontiers of Science Symposium and the Annual Meeting of the Society for the Scientific Study of Reading.

Correspondence concerning this article should be addressed to Shayne B. Piasta, Department of Teaching and Learning, The Ohio State University, Columbus, OH 43210. Phone: 614-688-4454; email: piasta.1@osu.edu

Abstract

Professional development (PD) is a potentially important mechanism for enhancing classroom practices and children's learning. In this large-scale randomized controlled trial, we examined the effectiveness of language and literacy PD, with and without coaching, offered at scale to early childhood educators (n = 546) across one state. Relative to the comparison condition, PD with coaching showed a small impact on the quantity of phonological awareness instruction, and PD with and without coaching impacted the quality of phonological awareness and writing instruction. PD did not impact children's (n = 1953, M_{age} = 4.53) emergent literacy skills, as measured by the research team, or kindergarten readiness, as measured by the state's kindergarten readiness assessment which exclusively focused on language and literacy skills. Although we can only speculate as to why this at-scale, state-sponsored PD did not realize intended impacts, these findings, as coupled with those from the literature, raise critical questions concerning current understandings of PD and the ability to achieve desired effects when implemented at scale.

Keywords: professional development, language and literacy, emergent literacy, kindergarten readiness

Educational Impact and Implications Statement

In the field of education, professional development is intended to improve classroom instruction and children's learning. However, we have a limited understanding as to its effects, especially when used at scale with large numbers of educators. In this study, we examined the language and literacy professional development offered to early childhood educators by one state. We found that the professional development affected only a few aspects of classroom literacy instruction and did not affect young children's literacy learning. These results suggest that, in order to be effective, at-scale professional development may require greater attention to design and implementation and highlight the need for pilot work regarding effects of professional development prior to large-scale investments.

At-Scale, State-Sponsored Language and Literacy Professional Development: Impacts on Early Childhood Classroom Practices and Children's Outcomes

Considerable research indicates the importance of young children's language and literacy skills for their future reading achievement. Both code-focused (e.g., alphabet knowledge, phonological awareness) and meaning-focused (e.g., oral language) skills are necessary for and predictive of children's later reading success (National Early Literacy Panel, 2008). A continuing need in the field is ensuring that young children have language- and literacy-learning experiences that optimize early learning in these areas; indeed, available evidence suggests that the quantity and quality of such classroom practices are highly variable in early education settings (Burchinal, Zaslow, & Tarullo, 2016; Pelatti, Piasta, Justice, & O'Connell, 2014). In addition to factors such as curriculum, classroom composition, staffing and resources, and relevant regulations (Early et al., 2010; Fuligni, Howes, Huang, Hong, & Lara-Cinisomo, 2012; LoCasale-Crouch et al., 2007), such variation may also be attributed to inconsistent education requirements for early childhood educators and variability in educator preparation programs (Early et al., 2007; Rhodes & Huston, 2012).

Currently, professional development (PD) is viewed as a potentially important mechanism for improving early childhood educators' language and literacy classroom practices and equalizing children's disproportionate access to high-quality early childhood education programs (Burchinal et al., 2016; Hamre, Partee, & Mulcahy, 2017; Rhodes & Huston, 2012). PD formats typically involve some element of group training through workshops or courses (e.g., Cunningham, Etter, Platas, Wheeler, & Campbell, 2015; Girolametto, Weitzman, Lefebvre, & Greenberg, 2007; Hamre et al., 2012), and may also include a coaching component (e.g., Buysse, Castro, & Peisner-Feinberg, 2010; Garet et al., 2008; Wasik & Hindman, 2011). Coaching refers

to ongoing, one-on-one mentoring of educators to support implementation of PD content, often working with educators within their own classrooms to model and observe practice. Regardless of format, the underlying logic model posits that PD will impact educators' knowledge, beliefs, and practices, and, ultimately, children's language and literacy skills (Desimone, 2009). Thus, two PD effects are inherently of interest: effects on educator outcomes and effects on child outcomes.

Effects of Language and Literacy PD on Early Childhood Educator and Child Outcomes

Findings within the PD literature indicate that PD, both with and without coaching, sometimes changes educator and/or child outcomes and other times does not result in intended changes (see Kraft, Blazar, & Hogan, 2018; Markussen-Brown et al., 2017 as recent meta-analyses). For example, Cunningham et al. (2015) examined educators' knowledge of and practices related to phonological awareness both prior to and following PD participation. Results indicated that educators entered the PD program with limited phonological awareness knowledge; correspondingly, phonological awareness instruction in these classrooms was limited (68% provided no phonological awareness instruction on a typical day) and judged to be of low quality per a standardized observation rating scale. Following the PD, educators' knowledge and the quantity and quality of classroom practices increased significantly, as did children's outcomes. Similarly, Wasik and Hindman (2011) conducted an experimental study in which educators participated in language and literacy PD with weekly coaching sessions. Results indicated that educators improved their quality of classroom practices, and children demonstrated significant improvement in two of three targeted domains.

In contrast, other studies have found that PD, both with and without coaching, did not have the intended impacts on educators or children. Girolametto et al. (2007) examined a PD

model that consisted of a 2-day workshop designed to facilitate language and print referencing use during shared storybook reading. Results indicated that educators increased their use of these strategies in only two of five categories; following PD, educators incorporated more inferential talk to connect storybook content to children's personal experiences and used more print references, but did not increase their extra-textual talk related to literal storybook elements, with respect to either individual objects/events or links among multiple objects/events, nor did they increase their extra-textual talk promoting analysis or evaluation of storybook events. Notably, children only demonstrated increases in response to educators' increased language and print referencing. Similarly, Buysse et al. (2010) examined changes in educators' instructional quality and children's skills among a large population of dual language learners. Educators participated in a 3-day PD series followed by alternating weeks of coaching and community meetings. Instruction and assessments were conducted in children's native language (i.e., English or Spanish). Results indicated that educators improved their classroom practice on three of six Early Language and Literacy Classroom Observation (ELLCO) subscales, namely the Literacy Activities Rating Scale of the original ELLCO (Education Development Center, 2002) and the addendums to the Classroom Observation Scale and Literacy Environment Checklist assessing practices for supporting dual language learners (Castro, 2005), and children improved on only one of five Spanish measures and no English measures. It is possible that these studies exhibited limited success relative to Cunningham et al. (2015) and Wasik and Hindman (2011) as they targeted broad skills rather than a narrow set of specific language and literacy skills, or, in the case of Girolametto et al. (2007), because the PD was of a shorter duration.

Several recent reviews and meta-analyses corroborate these findings by showing that educators might exhibit moderate increases in their knowledge, practice, and skills following PD

participation (Fukkink & Lont, 2007; Kraft et al., 2018; Markussen-Brown et al., 2017; Solomon, Klein, & Politylo, 2012), but children often make small gains, if any at all. These findings are generally reflective of the PD literature and suggest that PD, with or without coaching, can be effective for changing educator's knowledge, practice, or beliefs, but the degree to which it is effective varies by study.

At-Scale Studies of Language and Literacy PD for Early Childhood Educators

It is important to note that the vast majority of PD studies to date have been conducted under highly controlled conditions, and there is a need to better understand the extent to which effects hold when implemented at scale. From a scientific standpoint, it is important to understand the effectiveness of educational practices "under circumstances that would typically prevail in the target context" (Institute of Education Sciences & National Science Foundation, 2013, p. 10). Such at-scale effectiveness studies, by definition, are less controlled and not necessarily aimed at discerning mechanisms through which impacts are achieved (Gottfredson et al., 2015). Rather, effectiveness studies speak to external validity by determining whether practices realize intended effects when implemented in real-world contexts under routine conditions. Specifically, it is currently unclear if at-scale studies of language and literacy PD would produce effects similar to those of highly controlled studies as there have been few investigations of this nature (cf. Assel, Landry, Swank, & Gunnewig, 2007; Hamre et al., 2012; Jackson et al., 2006; Landry, Anthony, Swank, & Monseque-Bailey, 2009; Landry, Swank, Anthony, & Assel, 2011) and, in general, large-scale effectiveness studies often show more limited effects (e.g., Bleses et al., 2018).

Knowing whether at-scale PD achieves its intended results when implemented under authentic conditions is thus important not only for translational science but also educational

policy. In early childhood education, PD is increasingly seen as a policy lever for improving classroom practices and child outcomes. Various state and federal policies include requirements concerning PD for early childhood educators (e.g., quality rating and improvement systems, Head Start regulations; Connors & Morris, 2015; U.S. Department of Health and Human Services, 2016), despite little evidence as to whether these policies achieve intended results. Policy decisions grounded in current research might be ineffective if the impacts of PD differ when implemented under more authentic conditions with less researcher control. Given that at-scale PD efforts are already underway and involve tremendous costs in terms of time, resources, and money, it is imperative that we evaluate such endeavors.

Within the emerging body of at-scale PD literature, researchers have examined multiple PD models to test the presumed effects on educator and child outcomes. These include coaching only models (Landry et al., 2011), models with and without coaching (Assel et al., 2007; Jackson et al., 2006; Landry et al., 2009), and models that also integrate other supports such as online resources (Pianta, Mashburn, Downer, Hamre, & Justice, 2008). In all instances, PD groups showed significantly higher quantity and/or quality of classroom language and literacy practices than control groups on at least some measures. However, most studies showed that coaching did not have effects above and beyond group training (Assel et al., 2007; Jackson et al., 2006). In those instances in which coaching did appear to have an additional effect, these were moderated by educator or program factors (e.g., educators' initial quality of practice; Landry et al., 2009).

Measures of children's outcomes following educators' PD participation also exhibited mixed results. Some studies showed equivalent outcomes following PD both with and without coaching (Jackson et al., 2006) whereas others found that improvements were moderated by educator or program characteristics (e.g., curriculum; Assel et al., 2007; Landry et al., 2009).

Landry et al. (2011) found that children whose educators participated in two years of a coaching program made significantly greater gains than their peers whose educators only participated for one year. Notably, despite no substantial differences between PD only and coaching models on educators' practices, Assel et al. (2007) did find effects of both models on children's outcomes. Thus, there is a continuing need to understand whether at-scale PD, with and without coaching, realizes intended effects on practices and children's outcomes. This is particularly important given that educational policies require early childhood educators to regularly attend PD and, as noted above, often involve substantial financial and time commitments.

The Present Study

The current study contributes to this body of work by examining the effects of one state's language and literacy PD for early childhood educators as implemented at scale. This PD had been offered annually, free of cost, to all early childhood educators across the state for over 10 years. It consisted of 30-hrs of face-to-face workshops, and, similar to other recent large-scale PD efforts (Assel et al., 2007; Garet et al., 2008; Jackson et al., 2006; Landry et al., 2009), offered a complementary coaching component. The PD was designed with input from faculty at institutions of higher education across the state, and the content and delivery format incorporated many recommendations for effective PD derived from the literature (see Borko, 2004; Desimone, 2009; Garet, Porter, Desimone, Birman, & Yoon, 2001; Hamre et al., 2017; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007); notably, many of these features align with tenets of adult learning theories (e.g., Knowles, 1984; Mezirow, 1991; Putnam & Borko, 2000) and general principles of how people learn (e.g., active engagement, integration with prior knowledge, modeling, reflection, feedback; see also National Academies of Sciences, 2018).

The PD was both content-focused, featuring research-based information concerning

young children's language and literacy development, and practice-based, emphasizing how classroom instructional practices can support such development. PD topics were introduced over multiple sessions and attended to educators' existing knowledge, beliefs, and practices; the latter served as a starting point for building and integrating new information and opportunity to differentiate PD content. New information was accompanied by examples, and new practices were modeled. Educators had many opportunities to actively engage in learning via discussions, trialing new practices, and role playing. Educators also integrated new ideas into practice by trialing PD-based strategies in their own classrooms, after which they engaged in critical reflection and received feedback. PD sessions concluded with summaries of newly introduced content, to help solidify learning. The coaching component was intended to increase PD intensity via ongoing, one-on-one monthly mentoring over the course of the academic year. Coaches provided additional, differentiated learning opportunities via observation, modeling, goal setting, and provision of feedback based on educators' integration of PD practices in their classrooms. Like other PD efforts, the aim was to improve educator outcomes along with children's corresponding language and literacy skills, such that children were better prepared for kindergarten.

In an initial analysis (Piasta et al., 2017), we examined whether the language and literacy PD, with or without coaching, resulted in changes in educators' knowledge, beliefs, or practices. For the latter, we focused on the physical classroom literacy environment (e.g., presence of books, availability of literacy materials) and global instructional quality, as measured by the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2006). We focused on these aspects of practice because the physical literacy environment was targeted in the PD and we therefore expected to see change, and the CLASS is of interest to the state as it is increasingly

used to evaluate the quality of early childhood programs. What we learned was that, although we had reason to believe that the PD model would affect our selected outcomes, it had no impact on educators' knowledge, beliefs, or these two aspects of practice, even when we considered educator and classroom characteristics as potential moderators (Piasta et al., 2017).

Despite the findings of this initial analysis, several reasons support continued examination of the PD with respect to impacts on other aspects of practice and child outcomes. First, it is possible that impacts on educator outcomes were not captured in our initial work, given that knowledge and beliefs were measured via self-report surveys and our measure of quality (CLASS) was not language and literacy specific. In the current study, we examined the quantity and quality of observed language and literacy practices, which may be more likely to show change than the measures examined previously (Garet, Heppen, Walters, Smith, & Yang, 2016). Our language- and literacy-specific measures are also similar to those showing effects in previous at-scale PD work (Assel et al., 2007; Landry et al., 2009, 2011). Second, some evidence suggests that changes in educator knowledge or beliefs may not be necessary to yield effects on child outcomes (Hamre et al., 2012; Markussen-Brown et al., 2017). Indeed, some argue that educators must first alter practices and observe improved outcomes before changing their beliefs (e.g., Guskey, 1986), and additional research supports attending to both educator and child outcomes, given that associations between educators' practices and child outcomes are less straightforward than one might expect (Burchinal et al., 2016; Markussen-Brown et al., 2017). Finally, the effect of the state-sponsored PD on children's language and literacy outcomes has not yet been examined. Given that this was the ultimate goal of the PD, coupled with a need to examine practice in more detail, it seems critical to examine impacts on child outcomes.

In the current study, we investigated the impacts of this at-scale, state-sponsored PD on

the language and literacy practices implemented in early childhood classrooms as well as on children's emergent literacy skills and state kindergarten readiness outcomes; notably the state assessment focused exclusively on the domains of language and literacy. We addressed three questions: (1) To what extent does state-sponsored PD, with and without coaching, impact the quantity and quality of language and literacy practices experienced by children in early childhood classrooms? (2) To what extent does state-sponsored PD, with and without coaching, impact children's emergent literacy outcomes at the end of the preschool year and the fall of the following academic year? (3) For those children who matriculated to kindergarten, to what extent does state-sponsored PD, with and without coaching, impact outcomes on the state kindergarten readiness assessment?

Method

The present study was approved by the University Institutional Review Board and involved data collected from four sequential cohorts of early childhood educators who participated in state-sponsored language and literacy PD between 2010 and 2015. The PD had been developed and implemented by one mid-western state's Department of Education and its contractor, the Early Childhood Quality Network (ecQ-net), for over 10 years (for additional details, see Piasta et al., 2017; Weber-Mayrer, Piasta, & Pelatti, 2015); the research team was not involved in its development or implementation. This and other PD opportunities are regularly advertised and provided at no cost to all early childhood educators statewide. All educators who registered for state PD and met eligibility criteria (held a position as a lead, co-lead, or assistant educator, directly taught preschool-aged children, agreed to random assignment to PD condition) were invited to participate in the study. Originally, 760 educators agreed to participate and were randomly assigned to condition (PD, PD plus coaching [PD+], or comparison). However, many

educators (n = 213; 28%) withdrew prior to the start of PD/study activities, mainly due to changes in employment or scheduling such that they no longer met eligibility criteria (e.g., no longer in a classroom serving preschool children, administration no longer afforded time to attend PD). This turnover is consistent with the 29-30% rate reported for the general early childhood community (National Association for the Education of Young Children, 2004; Rhodes & Huston, 2012). Based on the overall attrition rate as well as attrition rates for each of the three study conditions [25%, 30%, and 29% for PD, PD+, and comparison respectively, which were not differential, χ^2 (2, N = 760) = 1.30, p = .522], the study meets the What Works Clearinghouse attrition standards (Institute of Education Sciences, 2017).

Participants

The analytic sample for the present study included 546 educators (one educator/ classroom had missing data for all outcomes), and data for analyses represented the 489 different early childhood classrooms in which these educators worked (53 classrooms had two participating educators, 2 classrooms had three participating educators). On average, educators were 42 years of age (SD = 10.88) and had 11 years of early childhood teaching experience (SD = 7.87). Most were female (92%; 7% unreported; note that total percentages may not equal 100% due to rounding) and not Hispanic or Latinx (81%; 18% unreported); 76% were White, 16% were Black, and 2% were multiracial or of other races (6% unreported). Educators' highest degrees included a high school diploma (16%), an Associate's degree (21%), a Bachelor's degree (29%), or graduate degree (23%; 11% unreported). The majority held some kind of state teaching certification (65%), and 14% held a Child Development Associate credential (10% unreported). Educators worked in rural (32%), suburban (29%), and urban (25%) settings (14% unreported). Most taught in classrooms situated within public schools (50%) or early childhood

centers (34%; 5% worked in home-based settings; 12% unreported) and held lead or co-lead positions (76%; 10% unreported). Twenty-four percent taught in early childhood special education settings (10% unreported), and 36% were affiliated with Head Start (5% unreported). Many reported using Creative Curriculum (48%); 26% used another commercially available curriculum but, with the exception of High/Scope (8%), none were used by more than 3% of educators (11% unreported). Class sizes averaged 18.67 children (*SD* = 7.78). Notably, educators did not statistically differ by condition on any reported characteristic.

Based on a priori power analyses, we sampled up to five children from the classrooms of each participating educator. Eligibility criteria included parental consent, at least 4 years of age, basic English proficiency, and no profound disabilities that prevented participation in study assessments. Children meeting eligibility criteria were classified into groups (expected to matriculate to kindergarten the following academic year, 5 years old but not matriculating to kindergarten, 4 years old but not matriculating to kindergarten), and research staff randomly selected children, prioritized by anticipated kindergarten matriculation and then age, until five (or all eligible) children were selected per classroom. In early childhood special education classrooms, research staff also considered special education status and randomly selected such that half of the sample from these classrooms were children with individual education plans (IEPs; alternating between selecting two or three children with IEPs per classroom) to approximate the inclusion model used in the state. The analytic sample included 1953 children enrolled in 482 different classrooms from whom we collected data on at least one assessment (no child data for 7 classrooms). On average, children were 4.53 years old (*SD* = 0.40) at the start of their study participation; 51% were boys (4% unreported), and 14% had an IEP (15% unreported). Most (71%) were White, 17% were Black, and 10% were multiracial or of other

races (3% unreported); 5% were Hispanic or Latinx. The highest degrees earned by children's mothers included high school diplomas or less (21%), high school diplomas plus additional training/certification (37%), Associate's degrees (13%), Bachelor's degrees (13%), or graduate degrees (12%; 4% unreported). For annual incomes, 40% of children's families reported $25,000 or less, 23% reported between $25,001 and $50,000, 12% reported between $50,001 and $75,000, and 21% reported more than $75,000 (5% unreported). Comparisons on child characteristics showed few differences across conditions, with a potentially different gender split [χ^2 (2, N = 1875) = 9.572, p = .008 but no standardized residuals exceeded 1.96] and more children who were Black in the comparison condition [χ^2 (4, N = 1903) = 13.994, p = .007]; thus, we included these as covariates in analyses. Importantly, children exhibited baseline equivalence on all measured language and literacy skills, with no significant differences across conditions and effect sizes (d) between 0 and 0.11 (Institute of Education Sciences, 2017).

Professional Development

A member of the research team randomly assigned participating educators to PD, PD+, or comparison conditions. Those assigned to PD or PD+ participated in 30 hrs of state-sponsored language and literacy training between September and January. This PD was offered by Early Language and Literacy Specialists (ELLS) employed by the state Department of Education; ELLS held at least a Master's degree in a relevant field. The PD consisted of ten 3-hr sessions delivered in a workshop format. The PD covered five major domains (two sessions each) related to supporting educators' use of language and literacy practices and children's emergent literacy outcomes (*Environment*, *Play*, *Oral Language*, *Early Reading*, and *Early Writing*) and was based on content derived from the research literature and the state early learning standards. PD sessions were organized into three parts (*Explorations*, *Implications and Demonstrations*, and

Connections to Teaching and Learning); these were strategically created in order to balance manualized content with acknowledgement of the existing knowledge, beliefs, and practices that participating educators brought to the PD. In *Explorations*, ELLS introduced educators to session topics and goals and explored educators' current knowledge and beliefs related to the session. In *Implications and Demonstrations*, ELLS provided both new content knowledge and new language and literacy practices; educators were provided with multiple examples, demonstrations, and opportunities to engage in supporting activities and to trial new practices. In *Connections to Teaching and Learning*, ELLS summarized session content, revisited session goals, and discussed ways to incorporate new content and practices into educators' classrooms. ELLS also explained the session's Into Practice activity, in which educators committed to implementing a new language and literacy practice in their classrooms, documented and reflected on this experience, and shared their experience in a subsequent PD session.

In addition to this 30-hr PD, educators randomly assigned to the PD+ condition also received ongoing individualized coaching throughout the academic year. Coaching was designed to align with the 30-hr workshops and further assist educators in integrating PD content pertaining to *Environment*, *Play*, *Oral Language*, *Early Reading*, and *Early Writing* into their classroom practices; educators were expected to receive 4-6 hrs of coaching per month. Coaches were experienced early childhood educators who volunteered for the state's coaching program as implemented by ecQ-net. Coaches received a coaching manual outlining the structure and content of the program, 24 hrs of coaching training, and ongoing support from designated ELLS. Coaching followed a cyclical model, in which educators worked with their coach to (a) determine a particular language and literacy practice to improve, (b) set goals for improvement, (c) plan for and implement the practice, (d) have their practice observed, (e) receive feedback on

the practice, and (f) engage in critical reflection. New goals were then set, and the cycle repeated. Educators documented their progress via portfolios, and coaches also completed online logs documenting coaching activities.

Educators randomly assigned to the comparison condition completed state-sponsored PD on alternative topics in order to provide a rigorous counterfactual for the study and decrease threats to validity (Shadish, Cook, & Campbell, 2002). In 2010 to 2011, educators in this condition selected and completed two 12-hr PD offerings on math, science, or social studies; these offerings followed a similar structure to the language and literacy PD (e.g. exploring educators' existing knowledge and beliefs about the topic, introducing new content knowledge and practices, summarizing and discussing ways to incorporate new content into classroom practices, and committing to complete an Into Practice activity) and were delivered using similar instructional strategies by contracted instructors. In 2012, these offerings were replaced by ecQ-net with a new 30-hr PD on young children's cognition, which integrated math, science, and social studies content into a single PD; educators in the comparison condition during 2012 to 2013 completed this PD offering which directly paralleled the format and delivery of the language and literacy PD. None of the comparison PD overlapped in content with the language and literacy PD.

PD fidelity. For purposes of the study, ecQ-net attempted to collect fidelity data concerning PD implementation. This included compiling PD attendance data from a state database and collecting survey data from ELLS/PD facilitators and participating educators. Survey response rates were low (less than 50% for ELLS/facilitators and less than 25% for educators), and survey data exhibited severe skew and limited variability. We thus provide fidelity data for descriptive purposes only, with the caveat that these data were reported by only a

subset of the sample. Attendance ranged from 80% to 100% (*M* = 98%, *SD* = 6%), and ELLS/facilitators rated participants' levels of participation at each session as high (*M* = 3.72, *SD* = 0.45 on a scale of 0 [low] to 4 [high]). Participating educators gave high ratings for the overall quality of the course and PD facilitator (*M* = 4.28, *SD* = 0.60 and *M* = 4.81, *SD* = 0.41, respectively; rated on scales of 1 [low] to 5 [high]). Using a checklist, participating educators also indicated that ELLS/facilitators implemented 68% to 100% of key components of PD sessions (*M* = 95%, *SD* = 7%). Coaching logs indicated that educators experienced 1 to 78 hrs of coaching (*M* = 28.62, *SD* = 19.99), with most coaching interactions focusing on the classroom literacy environment, supporting early reading, or administrative tasks (see Schachter, Weber-Mayrer, Piasta, & O'Connell, 2018).

Data Collection and Measures

Key data for the current study included the quantity and quality of language and literacy practices in participating classrooms, children's emergent literacy skills, and children's state kindergarten readiness assessment outcomes. We coded the quantity and quality of language and literacy practices from videotaped classroom observations conducted in the spring of the year in which educators participated in PD. We conducted direct assessments of children's emergent literacy skills in the fall and spring of the year during which their classroom educators participated in the PD as well as in the fall of the following academic year (fall follow-up). Research staff administered these assessments individually in a quiet location at children's early childhood education programs. Finally, we obtained state kindergarten readiness data for those children who matriculated to public kindergarten at the time of fall follow-up.

Quantity and quality of classroom practices. Educators participated in a classroom observation on a day considered representative of typical practice. Representativeness was

confirmed at the end of the observation by asking educators to rate "How typical of a day was today?" on a scale of 1 (not typical at all) to 5 (very typical; *M* = 4.22, *SD* = 1.02). Observations were videotaped and lasted the entirety of the classroom instructional day, as defined by the educator (*M* = 80 min, *SD* = 20 min). We assessed the *quantity of language and literacy practices* using an adapted version of the Individualizing Student Instruction coding scheme (see Connor et al., 2009; Pelatti et al., 2014 for additional details concerning the coding scheme and our observation and coding processes). This coding scheme exhaustively captured the amount of time, in min:sec, that each of the five participating children per classroom spent in activities related to any of the following language and literacy domains: Oral Language/Discussion, Vocabulary, Alphabet Knowledge, Print and Text Concepts, Word Identification, Phonological Awareness, Text Reading, Comprehension, and Writing. For purposes of the present study, we used these nine original language and literacy domains to compute quantity scores that mirrored those included in our quality measure (i.e., Oral Language = sum of Oral Language/Discussion and Vocabulary; Print and Letter Knowledge = sum of Alphabet Knowledge, Print and Text Concepts, and Word Identification; Book Reading = sum of Text Reading and Comprehension; Phonological Awareness and Writing remained their own domains; Overall = sum across all domains); intraclass correlations (ICCs) ranged from .81 to .99 across these domains for the approximately 20% of observations that we randomly selected for double-coding. We then averaged across the child-level scores within classrooms to create domain and overall scores for each classroom; if fewer than five children were participating in a given classroom, we randomly selected and coded the activities for additional children such that classroom scores were based on the experiences of five children. We assessed the *quality of language and literacy practices* using five language and literacy subscales of the Teacher Behavior Rating Scale (Oral Language,

Print and Letter Knowledge, Book Reading, Phonological Awareness, and Writing; Assel, Landry, & Swank, 2008). Each subscale includes multiple items representing instructional practices that support young children's language and literacy development; items are rated on a 4-point quality scale, with higher scores representing higher quality practices, and ICCs ranged .86 to .98 across these subscale scores for the 20% of observations randomly selected for double-coding. Item scores were averaged to derive the subscale quality scores used in analyses, and subscale scores were also averaged to create an overall quality score for each classroom. We also assessed the *global instructional quality* of classrooms using the Instructional Support domain of the CLASS, Pre-K version (Pianta et al., 2006) to serve as a covariate in analyses. The Instructional Support domain of CLASS indicates the quality of the general features of teacher-child instructional interactions on a 7-point scale. As recommended for coding CLASS from videotaped observations, each video was parsed into 20 min segments (Mashburn, Meyer, Allen, & Pianta, 2014). We randomly selected and coded three segments following standard CLASS procedures; these scores were averaged to create the score used in analyses. Double coding of 20% of segments indicated high coder agreement (88%).

Children's emergent literacy skills. Four aspects of children's emergent literacy skills were assessed at fall, spring, and fall follow-up. We assessed children's *oral language* via the core subtests of the Clinical Evaluation of Language Fundamentals Preschool-2 (internal consistencies range .82 to .83; Wiig, Secord, & Semel, 2004). For the Expressive Vocabulary subtest, children are asked to name objects, actions, and people, with items (20 max) scored on a scale of 0 to 2. For the Sentence Structure subtest, children are asked to indicate their understanding of sentences of increasingly complex syntax, with items (22 max) scored on a scale of 0 to 1. For the Word Structure subtest, children are asked to use increasingly complex

syntactical constructions, including pronouns, noun forms, verb tenses, and prepositions, with items (24 max) scored on a scale of 0 to1. To create the composite language score used in analyses, Expressive Vocabulary item scores were linearly transformed to a 0 to 1 scale to create a revised total score for this subtest before summing across the three subtests (Cronbach's α = .85 in the current sample); parallel analyses and exploratory factor analysis confirmed that the composite score measured a single construct. We assessed children's *alphabet knowledge* with respect to both letter-name knowledge and letter-sound knowledge. We assessed the former using the Uppercase and Lowercase Letter Recognition subtests of the Phonological Awareness Literacy Screening for Preschool (internal consistencies range .77 to .93; Invernizzi, Sullivan, Meier, & Swank, 2004). Children are presented with all 26 letters in a fixed, random order, first in uppercase and then in lowercase, and asked to name each letter; one point is awarded for each uppercase (0 to 26) and lowercase (0 to 26) letter correctly named. These scores were summed for the composite used in the current analyses (Cronbach's α = .98). For cohorts 3 and 4 (*n* = 998), we also assessed children's letter-sound knowledge using Letter Sound Short Forms (item response theory [IRT]-derived reliabilities ranged from .90 to .93 across 4 forms; Piasta, Phillips, Williams, Bowles, & Anthony, 2016; Cronbach's α = .83 to .89 across forms in the current sample). Children are presented with one randomly selected short form, consisting of six letters presented in both uppercase and lowercase, and asked to give the corresponding sound. All forms are equated in terms of difficulty. One point is awarded for each letter for which a correct sound is given, counting either short or long vowel sounds and any associated consonant sound as correct; these are summed, converted to IRT-based theta scores, and transformed to a scale ranging from 0 to 26. We assessed children's *phonological awareness* using two subtests of the Pre-Reading Inventory of Phonological Awareness (internal consistency for each subscale = .82;

Dodd, Cosbie, McIntosh, Teitzzel, & Ozanne, 2003). For the Rhyme Awareness subtest, children are presented with four words and asked to select the one that does not rhyme (12 items). For the Alliteration Awareness subtest, children are presented with four words and asked to select the one that does not have the same initial sound (12 items). Correct responses were tallied and summed across the two subtests to create the composite used in the current analyses (Cronbach's $\alpha = .75$); parallel analysis and exploratory factor analysis confirmed that the composite measured a single construct. We assessed children's *print concept knowledge* using the Preschool Print and Word Awareness assessment (IRT-derived reliability = .74; Justice, Bowles, & Skibbe, 2006; Cronbach's $\alpha = .73$ in the current sample). Children are asked questions eliciting their knowledge of 14 print concepts (12 items scored on scales of 0 to 1 or 0 to 2) within the context of a shared storybook reading. Scores were summed and converted to IRT-based scaled scores with a mean of 100 and standard deviation of 15.

Children's state kindergarten readiness outcomes. For children in cohorts 1, 2, and 3 who matriculated to public kindergarten at fall follow-up ($n = 605$), we obtained their scores on the state kindergarten readiness assessment, which focused exclusively on language and literacy skills. Kindergarten readiness data were not available for cohort 4 as the state implemented a new readiness assessment and declined to provide data. Items included answering *when* and *why* questions, repeating sentences, identifying letters, rhyming, and alliteration; scores could range from 0 to 29. Children's kindergarten teachers, who were blind as to children's assigned study condition, administered the assessment between the start of the academic year and November 1. The assessment technical report shows internal consistency of .84 and adequate item-fit statistics (American Institutes for Research, 2004).

Results

Descriptive statistics are provided in Table 1. Our analytic approach involved multiple steps. First, we conducted preliminary analyses to investigate the distributions of outcome variables of interest and patterns of missing data. This step informed subsequent analyses, including the selection of appropriate statistical procedures based on distributional assumptions and handling of missing data. Second, we used regression techniques to analyze the impacts of the PD and PD+ conditions relative to the comparison condition on the quantity and quality of educators' classroom language and literacy practices. In these analyses, our outcomes comprised the overall quantity and quality of language and literacy practices as well as the quantity and quality of practices related to each specific domain (Oral Language, Print and Letter Knowledge, Book Reading, Phonological Awareness, and Writing). Third, we used multilevel modeling to analyze the impacts of PD and PD+ relative to the comparison on children's language and literacy outcomes at spring and fall follow-up. Outcomes included children's scores on oral language, letter recognition, letter-sound knowledge, phonological awareness, and print concept knowledge measures at both time points and state kindergarten readiness scores at the fall follow-up time point for those who had matriculated to kindergarten. Due to the number of comparisons made, we applied the linear step-up procedure to control the familywise error rate (Benjamini & Hochberg, 1995).

Prior to beginning analyses, we carefully considered study condition assignment as this might affect results. Because educators, and not classrooms, were randomly assigned to study condition, a small percentage of classrooms (9%) had participating educators assigned to different conditions. We therefore coded condition in two ways: in the first, we assigned these classrooms to the most intensive language and literacy PD condition attended by any educator in that classroom (i.e., to PD+ if any educator in a given classroom participated in that condition, to

PD if any educator in a given classroom participated in that condition but none participated in PD+); in the second, we assigned these classrooms to the least intensive language and literacy PD condition attended by any educator in that classroom (i.e., to the comparison condition if any educator participated in that condition, to PD if any educator participated in the PD condition but no educators participated in the comparison). We conducted analyses for both ways of coding condition. The patterns of results were similar regardless of how condition was coded with one exception, and, for the sake of parsimony, we generally present only the results for classrooms when assigned to the most intensive PD condition. We describe the one exception in text and also describe the results of a post-hoc analysis in which we limited the sample to only those 434 classrooms with one participating educator (i.e., removed those classrooms with two or three participating educators who could have been assigned to different conditions); full results for all analyses are available from the first author upon request.

Preliminary Analyses

In preliminary analyses, we investigated the distributions of all outcome variables and patterns of missing data. All *quantity of classroom language and literacy practice* variables had non-normal, positively skewed distributions. For the *quality of language and literacy practice* variables, three were normally distributed (Overall, Oral Language, Book Reading), and three were non-normally distributed (Print and Letter Knowledge, Phonological Awareness, Writing; e.g., see Figure 1). The percentage of missing data was 8% for all classroom practice variables with the exception of the quality of Book Reading (38%); for this, many educators were missing values because they did not conduct a shared book reading activity during the observation. Variables pertaining to child emergent literacy skills and kindergarten readiness could be considered normally distributed; percentages of missing data ranged from 0.4% to 23%.

Separate variance *t*-tests indicated a systematic association between missingness on outcome variables and other variables in the dataset, suggesting a data pattern consistent with missing at random. Thus, as recommended by Graham (2012), we addressed the issue of missing data via multiple imputation. To accommodate the non-normal distributions, we used predictive mean matching in Stata v14 to impute and analyze the classroom outcome data, and we used Mplus v7.4 to impute and analyze the multilevel data for child outcomes. Note that for measures collected only for specific cohorts of children (i.e., letter-sound knowledge, kindergarten readiness), we only imputed data for children for whom we attempted to collect these data. For all, we created 40 imputed datasets.

Impact on Quantity and Quality of Classroom Language and Literacy Practices

To address our first research question, we compared classrooms on the overall quantity and quality of language and literacy practices provided as well as the quantity and quality of practices related to each specific domain. For the quantity variables, we made these comparisons using negative binomial regression given that the quantity data represented counts (i.e., number of minutes) and were positively skewed. For the quality variables, we used regression for outcomes that were normally distributed. For outcomes that were not normally distributed, we used median regression as a semiparametric approach that makes no assumptions about parametric or residual distributions (Cameron & Trivedi, 2009). We also confirmed these results using non-parametric analyses (available from the first author upon request).

Results are presented in Table 2. There were no significant differences in the quantity of language and literacy practices provided in either PD or PD+ classrooms versus comparison classrooms with the exception of the quantity of phonological awareness instructional practices. More phonological awareness instruction was provided in PD+ classrooms than comparison

classrooms, with a difference of approximately 0.47 min (i.e., 28 seconds). For the quality of language and literacy practices, analyses indicated some differences between PD or PD+ versus comparison classrooms. The quality of phonological awareness instruction in PD and PD+ classrooms was higher than in comparison classrooms when condition was based on the most intensive PD condition (see Figure 1). These differences were not significant when condition was based on the least intensive PD condition (coefficients = 0.127 and 0.152, *p*s = .483 and .374 for PD and PD+ versus comparison, respectively) but were significant and very similar in magnitude when utilizing the 434 classrooms with only one participating educator (both coefficients = 0.327, *p*s = .003 and .002). The quality of writing instruction was significantly higher in PD and PD+ than comparison classrooms (see Figure 1). There were no other differences between PD and PD+ versus comparison conditions in the quality of classroom practices.

Impact on Child Emergent Literacy and State Kindergarten Readiness Outcomes

To address our second and third research questions, we compared children's outcomes at the spring and fall follow-up timepoints using multilevel models to nest children within classrooms. ICCs for unconditional models ranged from .16 (phonological awareness) to .29 (print concept knowledge) for spring outcomes and .16 (letter-name knowledge) to .25 (language) for fall follow-up outcomes; the unconditional ICC for kindergarten readiness was .18. We fit separate multilevel models for each outcome with children's fall score(s), age, gender, race (recoded as White versus non-White), and maternal education level as covariates at level-1. Level-2 included the dummy-coded condition variables and classroom global instructional quality as covariates; for emergent literacy outcomes, we also included the relevant classroom-level mean of children's fall scores. Results contrasting spring and fall follow-up

outcomes for children in PD or PD+ conditions to those in the comparison condition are presented in Table 3 and Table 4, respectively. All results were the same regardless of how condition was coded. Children's language and literacy outcomes were not significantly different, nor did children differ in their state kindergarten readiness scores. Effect sizes (*d*) ranged from -0.15 to 0.10.

Discussion

The purpose of this study was to investigate the impacts of at-scale, state-sponsored PD, with and without coaching, on early childhood educators' language and literacy classroom practices and children's outcomes. We examined such impacts within the context of a rigorous effectiveness study, with the PD implemented under real-world conditions (Institute of Education Sciences & National Science Foundation, 2013). An affordance of this approach is that we were able to determine whether the state-sponsored PD, as authentically implemented, affected intended outcomes; by design, we were also able to determine the effects of PD with coaching and PD without coaching relative to the comparison condition. Beyond this, however, we are not able to make causal claims regarding the specific aspects of the PD that resulted in impacts or lack thereof as examining these mechanisms was not a goal of the current study.

Although we found some minimal changes in educators' practices as a result of PD participation, we did not detect impacts of PD on children's outcomes. These results are important to consider, within the context of state-sponsored PD as well as the larger early childhood community, given that the model was developed following research-based recommendations for effective PD (Borko, 2004; Desimone, 2009; Garet et al., 2001; Yoon et al., 2007) and the time and financial investments currently being made in early childhood PD nationally and internationally (Hamre et al., 2017; Vandenbroeck, Urban, & Peeters, 2016).

Below, we briefly discuss our main findings, before turning to broader implications.

Effects on Educators' Language and Literacy Practices

In the present study, we found impacts on two domains of educators' classroom practices using finer-grained observational measures of language and literacy instruction. These measures were closely aligned with the focus of the PD and allowed us to detect effects on educator outcomes not observed in our previous analysis, which utilized more global measures. Specifically, PD impacted both the quantity and quality of educators' phonological awareness instruction as well as the quality of writing instruction. Although these effects were small, it is notable that these were detected for areas of practice in which educators tend to provide minimal instruction and have a demonstrated need for additional support (Cunningham et al., 2015; Gerde, Bingham, & Pendergast, 2015). Moreover, the small effects are similar to those of other studies. For example, in a recent two-year study of the effects of PD with coaching, Bowne and colleagues (2016) found an average increase of 2.87 min in the amount of time that educators offered vocabulary support during a literacy lesson. Given that educators were not providing any vocabulary support prior to the PD, this seemingly small change translated into practical differences in practice. This may be similarly true for our participants, many of whom provided limited phonological awareness and writing instruction (see also Pelatti et al., 2014), although this remains an open question. Notably, our practice outcomes were based on one-day observations due to their time- and cost-intensive nature; although not atypical of the literature (e.g., Buysse et al., 2010; Cunningham et al., 2015; Jackson et al., 2006; Landry et al., 2009; Landry et al., 2011; Wasik & Hindman, 2011), we acknowledge that multi-day observations might yield more generalizable results.

With particular respect to the effects of coaching, our results suggest minimal additional

benefits for educators' classroom practices. Specifically, PD with coaching impacted the quantity of phonological awareness instruction whereas PD with and without coaching both impacted the quality of phonological awareness and writing instruction (and both conditions also had similar means and medians for these outcomes). These findings are consistent with the equivocal results in prior work contrasting coaching and other PD formats (e.g., Assel et al., 2007; Jackson et al., 2006; Lonigan, Farver, Phillips, & Clancy-Menchetti, 2011) as well as other research showing that coaching sessions may only be effective when specific strategies are incorporated (e.g., performance feedback; Noell et al., 2005, 2014). Our results are qualified by how coaching was implemented within this at-scale, state-sponsored PD: Coaching was provided by peer early childhood educators, as opposed to experts or highly trained coaches (Hamre et al., 2017; Sanetti & Kratochwill, 2007) and, although most educators (80%) experienced at least 10 hrs of coaching, the amount of coaching varied from 1 to 78 hrs. Only 19% of educators experienced the intended 48 hrs of coaching, and further analysis indicated that some of this time was spent on administrative tasks (e.g., scheduling, paperwork; Schachter et al., 2018). Such less-than-ideal implementation is perhaps unsurprising in a statewide coaching program but may explain the lack of effects in the current study as well as recent meta-analytic results showing that the effects of coaching are greatly attenuated when implemented at scale (Kraft et al., 2018). More work is necessary to identify key features of coaching that promote educator outcomes and can help explain the mixed findings in the literature (Schachter, 2015; Solomon et al., 2012). It remains unclear, for example, how the amount or intensity of coaching is related to educator outcomes (Weber-Mayrer, Piasta, Ottley, Justice, & O'Connell, 2018). Such continued work is particularly necessary given steady increases in large-scale use of coaching (e.g., recent Head Start initiatives to include coaching in all programs; U.S. Department

of Health and Human Services, 2016), to inform how such PD models can be supported at scale.

Effects on Child Outcomes

With respect to child language and literacy outcomes, we were unable to detect any impacts of the PD at the end of the preschool year or at fall follow-up. We carefully selected valid and reliable measures that are widely used in the literature and were aligned with the content of the language and literacy PD. However, it is possible that our measures were insufficient for measuring such impacts. For example, our measure of phonological awareness assessed children's abilities related to rhyming and initial sound awareness; it is possible that educators targeted other phonological awareness skills (e.g., segmenting or blending of syllables or compound words) and that measuring such skills may have shown effects. Similarly, although the PD impacted the quality of writing instruction, we did not analyze corresponding effects on children's emergent writing skills. Notably, we also did not detect any impact on children's kindergarten readiness scores, as measured by a state assessment that emphasized the language and literacy domains targeted in the PD. The latter is important given that it comprises an authentic measure utilized by policymakers and practitioners and that improving kindergarten readiness constituted one impetus for the state's investment in PD for early childhood educators.

Our findings parallel other PD studies reporting impacts on educator outcomes but not child outcomes (e.g., Gerde, Duke, Moses, Spybrook, & Shedd, 2014; Yoshikawa et al., 2015), particularly those that focused on a broad range of language and literacy skills (e.g., Buysse et al., 2010; Girolametto et al., 2007), albeit within the context of this at-scale, state-implemented PD. Whereas these null findings are disappointing, they are not surprising given the minimal impacts on educators' knowledge, beliefs, and practices as demonstrated in this and our previous analysis (Piasta et al., 2017). Drawing on the logic model underlying this and other PD efforts,

we may need more substantial impacts on educator outcomes to yield effects at the child level; one recent analysis suggests that a 1 *SD* change in educator practice is required for a corresponding 0.2 *SD* change in child outcomes (Kraft et al., 2018). Alternatively, it may be that effects on child outcomes are not immediate; perhaps it requires multiple years of educators implementing PD-supported practices in their classrooms to refine these practices, integrate these with existing practices, and observe changes in children's learning (e.g., Landry et al., 2011).

Implications for Early Childhood Professional Development

Our findings, when interpreted within the context of the extant research literature, raise important questions regarding the field's approach to PD that have not yet been addressed. Addressing these gaps has important implications for future research and efforts to support professional learning as a means of promoting both educator and child outcomes.

Current recommendations for effective PD converge on several key principles, such as the need for PD to be embedded, intensive, and ongoing, exhibit a content focus, provide active learning opportunities, and be responsive to educators' needs (Borko, 2004; Desimone, 2009; Garet et al., 2001; Yoon et al., 2007). The current state-sponsored PD was designed to reflect these recommendations, although we acknowledge that the PD, as implemented, may not have fully adhered to these principles or perhaps did not emphasize the most important principles. Indeed, limitations of this work include the fact that the PD was not initially subjected to highly controlled efficacy trials and that the current study was designed to evaluate the PD as a whole and as typically implemented. Thus, we are unable to disentangle whether the minimal effects were due to the design, content, and/or format of the PD itself or challenges of implementing the PD at scale; the limited variability and low response rate for fidelity measures also constrain our ability to disentangle results. Yet, our findings and the collective PD literature suggest that

adhering to these general principles may not be sufficient to realize intended effects, and findings from other literatures suggest additional elements necessary to support educator change (e.g., performance feedback, directed rehearsal, self-monitoring and evaluation; Hagermoser Sanetti, Collier-Meek, Long, Byron, & Kratochwill, 2015; Noell et al., 2014; Sanetti & Kratochwill, 2007; Solomon et al., 2012). More research is crucial to understand how best to integrate and enact these principles in practice, particularly when PD is to be implemented at scale (Hamre et al., 2017; Schachter, 2015). For instance, coaching continues to be a key PD format because of its potential to reflect these key principles, especially in its affordances for PD that is embedded and differentiated. The mixed evidence for coaching, however, suggests that the specific ways that coaching is implemented to meet these principles matters (e.g., Noell et al., 2005; Weber-Mayrer et al., 2018). More generally, we continue to know little about how to adapt PD to better respond to individual educators' needs (Borko, 2004). By necessity, the current PD served a broad, statewide population of early childhood educators, with greatly varying backgrounds, knowledge, beliefs, and skills (Weber-Mayrer et al., 2015). Although it included opportunities to tailor content to participants' needs (e.g., *Explorations* component, coaching), we are unable to discern whether such adjustments improved the PD or perhaps altered the intended focus or intensity of the PD (Schachter et al., 2018). More work on how to balance the integrity of PD with ways of increasing responsivity may be crucial for implementing effective PD at scale.

Similarly, parameters concerning the intensity and duration of PD also require further study. Although some research indicates that as little as 10 hrs of contact time may affect outcomes (Gerde et al., 2014; cf. Yoon et al., 2007), it is likely that the necessary intensity and duration depends on the complexity of the knowledge and practices targeted by the PD. Moreover, research has demonstrated that when educators participated in PD with a longer

duration (i.e., 2 years), children had significantly better outcomes (Landry et al., 2011). Observing impacts on educator outcomes may similarly require longer periods of time, as educators balance the integration of new PD content with the many demands that they face in their day-to-day instruction (Schachter, 2015). More research is needed concerning intensity and duration as these relate to targets of PD, as well as the time frame during which intended effects may be observed. This includes research regarding the necessary infrastructure, resources, and other supports to achieve sufficient levels of intensity and duration in large-scale PD.

Another open question concerns understandings as to how PD brings about change and, in particular, relations between effects of PD on educator and child outcomes. Prevailing logic models (e.g., Desimone, 2009) posit that PD will affect knowledge and beliefs, thereby changing practice and, ultimately, improving child outcomes. Under this model, a certain level of impacts on educator outcomes may be necessary to achieve impacts on child outcomes. Given that at-scale implementation typically results in smaller effects than highly controlled studies (Bleses et al., 2018; Kraft et al., 2018), investigating thresholds for achieving intended PD impacts may be worthwhile. Moreover, Burchinal and colleagues (2016) found that additional increases in the quality of classroom practices only promoted children's outcomes when a certain threshold of quality was already established and that aspects of practice and child outcomes were related in non-linear ways. It is also possible that this logic model is not accurate (Hamre et al., 2012). New research suggests that associations among educators' knowledge, beliefs, and practice may be more complex than acknowledged in this model (Schachter, Spear, Piasta, Justice, & Logan, 2016), and others have theorized that change in educators' knowledge or beliefs may occur only after extended opportunities to implement new practices or observe child-level change (Guskey, 1986; Lieber et al., 2010; Pianta et al., 2014). Thus, the linear relations posited in prevailing

logic models of PD may need to be reexamined as this research base grows, with continued attention to the mechanisms that can support meaningful educator and child impacts.

Altogether, critical questions remain concerning the underlying logic of and effective approaches to PD. In the present study, the at-scale, state-sponsored PD did not achieve desired effects on child outcomes despite some effects on educators' language and literacy practices. Rather, results reveal the need for continued research on PD to avoid potentially premature and costly efforts of at-scale implementation and to realize the intended benefits of investing in early childhood education for optimizing children's learning and development.

References

American Institutes for Research. (2004, April). Kindergarten entry-level literacy instrument field-test technical report. Columbus, OH: Author.

Assel, M. A., Landry, S. H., & Swank, P. R. (2008). Are early childhood classrooms preparing children to be school ready?: The CIRCLE Teacher Behavior Rating Scale. In L. M. Justice & C. Vukelich (Eds.), *Achieving excellence in preschool literacy instruction* (pp. 120-135). New York: Guilford Press.

Assel, M. A., Landry, S. H., Swank, P. R., & Gunnewig, S. (2007). An evaluation of curriculum, setting, and mentoring on the performance of children enrolled in pre-kindergarten. *Reading and Writing, 20*, 463-494.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*, 289-300.

Bleses, D., Højen, A., Justice, L. M., Dale, P. S., Dybdal, L., Piasta, S. B., . . . Haghish, E. F. (2018). The effectiveness of a large-scale language and preliteracy intervention: The spell randomized controlled trial in Denmark. *Child Development, 89*, 342-363. doi: 10.1111/cdev.12859

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*, 3-15. doi: 10.3102/0013189x033008003

Bowne, J. B., Yoshikawa, H., & Snow, C. E. (2016). Experimental impacts of a teacher professional development program in early childhood on explicit vocabulary instruction across the curriculum. *Early Childhood Research Quarterly, 34*, 27-39.

Burchinal, M. R., Zaslow, M. J., & Tarullo, L. (Eds.). (2016). *Quality thresholds, features, and dosage in early care and education: Secondary data analyses of child outcomes*. Washington, DC: Society for Research in Child Development.

Buysse, V., Castro, D. C., & Peisner-Feinberg, E. (2010). Effects of a professional development program on classroom practices and outcomes for Latino dual language learners. *Early Childhood Research Quarterly, 25*, 194-206.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconomics using stata* (Rev. ed.). College Station, TX: STATA Press.

Castro, D. C. (2005). *Early language and literacy classroom observation addendum for English language learners*. Chapel Hill, NC: University of North Carolina, FPG Child Development Institute.

Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., . . . Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher, 38*, 85-99. doi: 10.3102/0013189X09332373

Connors, M. C., & Morris, P. A. (2015). Comparing state policy approaches to early care and education quality: A multidimensional assessment of quality rating and improvement systems and child care licensing regulations. *Early Childhood Research Quarterly, 30*, 266-279. doi: 10.1016/j.ecresq.2014.05.006

Cunningham, A. E., Etter, K., Platas, L., Wheeler, S., & Campbell, K. (2015). Professional development in emergent literacy: A design experiment of teacher study groups. *Early Childhood Research Quarterly, 31*, 62-77. doi: 10.1016/j.ecresq.2014.12.002

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181-199. doi: 10.3102/0013189x08331140

Dodd, B., Cosbie, S., McIntosh, B., Teitzzel, T., & Ozanne, A. (2003). *Pre-reading inventory of phonological awareness*. San Antonio, TX: PsychCorp.

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*, 177-193. doi: 10.1016/j.ecresq.2009.10.003

Early, D. M., Maxwell, K. L., Burchinal, M. R., Bender, R. H., Ebanks, C., Henry, G. T., . . . Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*, 558-580. doi: 10.1111/j.1467-8624.2007.01014.x

Education Development Center. (2002). *Early language and literacy observation toolkit*. Baltimore, MD: Paul H. Brookes.

Fukkink, R. G., & Lont, A. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly, 22*, 294-311.

Fuligni, A. S., Howes, C., Huang, Y., Hong, S. S., & Lara-Cinisomo, S. (2012). Activity settings and daily routines in preschool classrooms: Diverse experiences in early learning settings for low-income children. *Early Childhood Research Quarterly, 27*, 198-209. doi: 10.1016/j.ecresq.2011.10.001

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., . . . Regional, A. (2008). The impact of two professional development interventions on early reading instruction

and achievement. Ncee 2008-4030: National Center for Education Evaluation and Regional Assistance.

Garet, M. S., Heppen, J., Walters, K., Smith, T., & Yang, R. (2016). Does content-focused teacher professional development work? Findings from three Institute of Education Sciences studies. Washington, D.C.: Institiute of Education Sciences, National Center for Educationa Evaluation and Regional Assistance.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*, 915-945. doi: 10.3102/00028312038004915

Gerde, H. K., Bingham, G. E., & Pendergast, M. L. (2015). Reliability and validity of the writing resources and interactions in teacher environments (write) for preschool classrooms. *Early Childhood Research Quarterly, 31*, 34-46.

Gerde, H. K., Duke, N. K., Moses, A. M., Spybrook, J., & Shedd, M. K. (2014). How much for whom? Lessons from an efficacy study of modest professional development for child care providers. *Early Education and Development, 25*, 421-441.

Girolametto, L., Weitzman, E., Lefebvre, P., & Greenberg, J. (2007). The effects of in-service education to promote emergent literacy in child care centers: A feasibility study. *Language, Speech, and Hearing Services in Schools, 38*, 72-83.

Gottfredson, D., Cook, T., Gardner, F. M., Gorman-Smith, D., Howe, G., Sandler, I., & Zafft, K. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16*, 898-926. doi: 10.1007/s11121-015-0555-x

Graham, J. W. (2012). *Missing data: Analysis and design*. New York, NY: Springer.

Guskey, T. R. (1986). Staff development and the process of teacher change. *Educational Researcher, 15*, 5-12.

Hagermoser Sanetti, L. M., Collier-Meek, M. A., Long, A. C. J., Byron, J., & Kratochwill, T. R. (2015). Increasing teacher treatment integrity of behavior support plans through consultation and implementation planning. *Journal of School Psychology, 53*, 209-229. doi: 10.1016/j.jsp.2015.03.002

Hamre, B. K., Partee, A., & Mulcahy, C. (2017). Enhancing the impact of professional development in the context of preschool expansion. *AERA Open, 3*, 2332858417733686. doi: 10.1177/2332858417733686

Hamre, B. K., Pianta, R. C., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J. T., . . . Scott-Little, C. (2012). A course on effective teacher-child interactions. *American Educational Research Journal, 49*, 88-123. doi: 10.3102/0002831211434596

Institute of Education Sciences. (2017, October). What Works Clearinghouse standards handbook (version 4.0). Retrieved August 17, 2018 https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Institute of Education Sciences, U. S. D. o. E., & National Science Foundation. (2013). *Common guidelines for education research and development*. Washington, DC: Authors.

Invernizzi, M. A., Sullivan, A., Meier, J. D., & Swank, L. (2004). *Phonological awareness literacy screening for preschool: Teacher's manual*. Charlottesville, VA: University of Virginia.

Jackson, B., Larzelere, R., Clair, L. S., Corr, M., Fichter, C., & Egertson, H. (2006). The impact of HeadsUp! Reading on early childhood educators' literacy practices and preschool

children's literacy skills. *Early Childhood Research Quarterly, 21*, 213-226. doi: 10.1016/j.ecresq.2006.04.005

Justice, L. M., Bowles, R. P., & Skibbe, L. E. (2006). Measuring preschool attainment of print-concept knowledge: A study of typical at-risk 3- to 5-year-old children using item response theory. *Language, Speech, and Hearing Services in Schools, 37*, 224-235.

Knowles, M. S. (1984). *Andragogy in action: Applying modern principles of adult education*. San Francisco, CA: Jossey Bass.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*, 547-588. doi: 10.3102/0034654318759268

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*, 448-465. doi: 10.1037/a0013842

Landry, S. H., Swank, P. R., Anthony, J. L., & Assel, M. A. (2011). An experimental study evaluating prefssional development activities within a state funded pre-kindergarten program. *Reading and Writing, 24*, 971-1010.

Lieber, J., Butera, G., Hanson, M., Palmer, S., Horn, E., & Czaja, C. (2010). Sustainability of a preschool curriculum: What encourages continued use among teachers? *NHSA Dialog: A Research-to-Practice Journal for the Early Childhood Field, 13*, 225-242.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., . . . Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22*, 3-17. doi: 10.1016/j.ecresq.2006.05.001

Lonigan, C. J., Farver, J., Phillips, B., & Clancy-Menchetti, J. (2011). Promoting the development of preschool children's emergent literacy skills: A randomized evaluation of a literacy-focused curriculum and two professional development models. *Reading and Writing, 24*, 305-337. doi: 10.1007/s11145-009-9214-6

Markussen-Brown, J., Juhl, C. B., Piasta, S. B., Bleses, D., Højen, A., & Justice, L. M. (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. doi: 10.1016/j.ecresq.2016.07.002

Mashburn, A. J., Meyer, J. P., Allen, J. P., & Pianta, R. C. (2014). The effect of observation length and presentation order on the reliability and validity of an observational measure of teaching quality. *Educational and Psychological Measurement, 74*, 400-422. doi: 10.1177/0013164413515882

Mezirow, J. (1991). *Transformative dimensions of adult learning*. San Francisco, CA: Jossey Bass.

National Academies of Sciences, E., and Medicine,. (2018). *How people learn ii: Learners, contexts, and cultures*. Washington, D.C.: The National Academies Press.

National Association for the Education of Young Children. (2004). NAEYC advocacy toolkit Retrieved from www.naeyc.org/files/naeyc/file/policy/toolkit.pdf

National Early Literacy Panel. (2008). *Developing early literacy*. Washington, D. C.: National Institute for Literacy.

Noell, G. H., Gansle, K. A., Mevers, J. L., Knox, R. M., Mintz, J. C., & Dahir, A. (2014). Improving treatment plan implementation in schools: A meta-analysis of single subject

design studies. *Journal of Behavioral Education, 23*, 168-191. doi: 10.1007/s10864-013-9177-1

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., . . . Duhon, G. J. (2005, 2005 Winter). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34,* 87-106.

Pelatti, C. Y., Piasta, S. B., Justice, L. M., & O'Connell, A. A. (2014). Language- and literacy-learning opportunities in early childhood classrooms: Children's typical experiences and within-classroom variability. *Early Childhood Research Quarterly, 29*, 445-456. doi: 10.1016/j.ecresq.2014.05.004

Pianta, R. C., Burchinal, M., Jamil, F. M., Sabol, T., Grimm, K., Hamre, B. K., . . . Howes, C. (2014). A cross-lag analysis of longitudinal associations between preschool teachers' instructional support identification skills and observed behavior. *Early Childhood Research Quarterly, 29*, 144-154. doi: 10.1016/j.ecresq.2013.11.006

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2006). *Classroom assessment scoring system: Pre-k*. Charlottesville, VA: Center for Advanced Study of Teaching and Learning.

Pianta, R. C., Mashburn, A. J., Downer, J. T., Hamre, B. K., & Justice, L. M. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*, 431-451.

Piasta, S. B., Justice, L. M., O'Connell, A. A., Mauck, S. A., Weber-Mayrer, M., Schachter, R. E., . . . Spear, C. F. (2017). Effectiveness of large-scale, state-sponsored language and literacy professional development on early childhood educator outcomes. *Journal of Research on Educational Effectiveness, 10*, 354-378.

Piasta, S. B., Phillips, B. M., Williams, J. M., Bowles, R. P., & Anthony, J. L. (2016). Measuring young children's alphabet knowledge: Development and validation of brief letter-sound knowledge assessments. *The Elementary School Journal, 116*, 523-548.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*, 4-15.

Rhodes, H., & Huston, A. (2012). Building the workforce our youngest children deserve. *Social Policy Report* (Vol. 26, pp. 3-26). Ann Arbor, MI: Society for Research in Child Development.

Sanetti, L. M. H., & Kratochwill, T. R. (2007). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral Consultation and Therapy, 4*, 95-114. doi: 10.1037/h0100835

Schachter, R. E. (2015). An analytic study of the professional development research in early childhood education. *Early Education and Development, 26*, 1057-1085. doi: 10.1080/10409289.2015.1009335

Schachter, R. E., Spear, C. F., Piasta, S. B., Justice, L. M., & Logan, J. A. R. (2016). Early childhood educators' knowledge, beliefs, education, experiences, and children's language- and literacy-learning opportunities: What is the connection? *Early Childhood Research Quarterly, 36*, 281-294. doi: 10.1016/j.ecresq.2016.01.008

Schachter, R. E., Weber-Mayrer, M. M., Piasta, S. B., & O'Connell, A. A. (2018). What do coaches and teachers talk about during coaching?: Coaches' reports of the content of coaching interactions. *Early Education and Development, 29*, 852-872.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.

Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review, 41*, 160-175.

U.S. Department of Health and Human Services. (2016). Head Start program performance standards. Washington, D.C.: Head Start Bureau.

Vandenbroeck, M., Urban, M., & Peeters, J. (Eds.). (2016). *Pathways to professionalism in early childhood education and care*. New York, NY: Routledge.

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*, 455-469. doi: 10.1037/a0023067

Weber-Mayrer, M. M., Piasta, S. B., Ottley, J. R., Justice, L. M., & O'Connell, A. A. (2018). Early childhood literacy coaching: An examination of coaching intensity and changes in educators' literacy knowledge and practice. *Teaching and Teacher Education, 76*, 14-24.

Weber-Mayrer, M. M., Piasta, S. B., & Pelatti, C. Y. (2015). State-sponsored professional development for early childhood educators: Who participates and associated implications for future offerings. *Journal of Early Childhood Teacher Education, 36*, 44-60. doi: 10.1080/10901027.2014.996927

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical evaluation of language fundamentals preschool* (2nd ed.). San Antonio, TX: Harcourt Assessment.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, Regional Educational Laboratory Southwest.

Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M. C., Weiland, C., . . . Arbour, M. C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology, 51*, 309-322. doi: 10.1037/a0038785

Table 1

Descriptive statistics by condition for classroom language and literacy practices, children's emergent literacy skills, and children's kindergarten readiness

	PD			PD+			Comparison		
	M (Mdn)	*SD*	Range	*M (Mdn)*	*SD*	Range	*M (Mdn)*	*SD*	Range
Quantity of language and literacy practices (spring; in min)									
Overall	13.77 (12.35)	8.84	0.15–39.77	14.08 (13.16)	8.70	0.00–46.86	12.93 (10.55)	10.16	0.00–58.92
Oral language	3.35 (1.71)	4.04	0.00–20.50	3.55 (2.17)	3.97	0.00–18.43	3.53 (2.21)	4.57	0.00–32.98
Print and letter knowledge	3.22 (2.11)	3.26	0.00–16.92	2.92 (1.69)	3.34	0.00–16.11	2.74 (1.34)	3.96	0.00–28.95
Book reading	5.92 (5.16)	5.32	0.00–28.71	6.26 (5.78)	5.17	0.00–26.62	5.66 (4.49)	5.86	0.00–37.73
Phonological awareness	0.62 (0.00)	1.60	0.00–9.39	0.79 (0.00)	2.82	0.00–26.14	0.26 (0.00)	0.93	0.00–6.49
Writing	0.62 (0.00)	1.57	0.00–11.35	0.50 (0.00)	1.34	0.00–9.24	0.72 (0.00)	2.73	0.00–25.35
Quality of language and literacy practices (spring; 0 to 3 rating scale)									
Overall	1.52 (1.55)	0.32	0.48–2.28	1.53 (1.56)	0.32	0.53–2.32	1.47 (1.48)	0.33	0.44–2.35
Oral language	1.80 (1.83)	0.35	0.88–2.67	1.81 (1.88)	0.36	0.75–2.75	1.80 (1.88)	0.36	0.88–2.63
Print and letter knowledge	1.69 (2.00)	0.47	0.00–2.50	1.64 (1.75)	0.49	0.00–3.00	1.63 (2.00)	0.55	0.00–3.00
Book reading	1.82 (1.86)	0.33	0.96–2.61	1.76 (1.82)	0.37	0.58–2.59	1.73 (1.79)	0.36	0.79–2.60
Phonological awareness	0.89 (1.00)	0.65	0.00–2.33	0.94 (1.00)	0.60	0.00–2.00	0.74 (0.67)	0.63	0.00–2.33
Writing	1.33 (1.50)	0.74	0.00–3.00	1.35 (1.50)	0.70	0.00–2.67	1.13 (1.00)	0.77	0.00–2.50
Children's emergent literacy skills									
Oral language (max of 66)									
Fall	41.06	9.82	1.00–61.00	40.04	9.90	13.00–61.00	39.68	9.75	3.00–65.00
Spring	47.05	9.07	10.50–66.00	45.57	9.30	10.00–64.00	46.90	9.48	1.00–66.00
Follow-up	51.59	8.19	19.00–66.00	50.97	8.70	7.50–66.00	51.15	9.32	14.50–66.00

	PD			PD+			Comparison		
	M (Mdn)	*SD*	Range	*M (Mdn)*	*SD*	Range	*M (Mdn)*	*SD*	Range
Letter recognition (max of 52)									
Fall	26.50	17.87	0–52	25.59	18.62	0–52	24.71	18.20	0–52
Spring	35.82	16.12	0–52	34.16	17.12	0–52	34.22	16.91	0–52
Follow-up	44.27	11.65	0–52	44.04	12.38	0–52	43.20	13.17	0–52
Letter sounds (max of 26)[a]									
Fall	9.19	8.40	1.03–23.53	8.55	8.03	1.03–23.53	8.59	8.00	1.03–23.53
Spring	13.02	8.88	1.03–23.53	11.71	8.48	1.03–23.53	12.62	8.58	1.03–23.53
Follow-up	16.92	8.14	1.03–23.53	18.26	7.03	1.03–23.53	17.45	7.42	1.03–23.53
Phonological awareness (max of 24)									
Fall	7.53	4.74	0–24	7.43	4.41	0–24	7.40	4.59	0–24
Spring	9.93	6.04	0–24	9.46	5.77	0–24	9.60	5.55	0–24
Follow-up	13.13	6.01	0–24	13.46	6.22	0–24	13.10	6.02	0–24
Print concept knowledge (max of 161)									
Fall	105.52	18.05	46–161	105.45	16.23	46–161	105.71	16.43	46–161
Spring	116.30	19.35	46–161	114.52	18.53	64–161	113.60	19.51	46–161
Follow-up	127.67	17.37	82–161	129.01	18.54	46–161	127.74	18.06	46–161
Children's kindergarten readiness (max of 29)[b]									
Follow-up	22.02	5.67	5–29	21.36	6.31	5–29	21.18	6.54	5–29

Note. Condition based on most intensive PD condition. Medians (*mdn*) presented for classroom practice variables given the non-normal distributions for many of these outcomes. No significant differences across conditions on children's fall emergent literacy skills (*d*s < 0.11).

[a]Letter sound assessment was administered only to cohorts 3 and 4 (n = 998). [b]Kindergarten readiness assessment data, which focused exclusively on language and literacy, was obtained only for children in cohorts 1, 2, and 3 who matriculated to public kindergarten (n = 605).

Table 2

Results for PD and PD+ versus comparison conditions on the quantity and quality of classroom language and literacy practices

		PD v. Comparison		PD+ v. Comparison	
	ICC	coeff	*p*	coeff	*p*
Quantity of language and literacy practices[a]					
Overall	.97	0.06	.467	0.08	.316
Oral language	.81	-0.06	.659	-0.00	.991
Print and letter knowledge	.93	0.16	.224	0.07	.610
Book reading	.99	0.05	.710	0.10	.421
Phonological awareness	.99	0.73	.031	0.95	.004*
Writing	.90	-0.13	.648	-0.33	.267
Quality of language and literacy practices					
Overall[b]	.94	0.05	.232	0.05	.159
Oral language[b]	.87	0.00	.947	0.01	.756
Print and letter knowledge[c]	.88	0.06	.673	-0.19	.158
Book reading[b]	.99	0.06	.233	0.02	.653
Phonological awareness[c]	.96	0.33	.002*	0.33	.001*
Writing[c]	.95	0.50	.011*	0.50	.009*

Note. Results based on full sample (n = 489 classrooms) and 40 imputed datasets. A priori alpha level set at .01. ICC = intraclass correlation as a measure of interrater reliability (as computed from double coded observations); coeff = coefficient. [a]Comparison via negative binomial regression analysis. [b]Comparison via regression analysis. [c]Comparison via median regression.

*Comparison was statistically different when applying the linear step-up procedure to control familywise error rate.

Table 3

Multilevel results for PD and PD+ versus comparison on children's spring emergent literacy skills

	Oral Language		Letter Recognition		Letter Sounds		Phonological Awareness		Print Concept Knowledge	
	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*p*
Unconditional ICC	.22		.19		.29		.16		.26	
Classroom level										
Intercept γ_{00}	45.17	<.001	33.31	<.001	11.65	<.001	8.66	<.001	110.75	<.001
PD γ_{01}	-0.37	.306	0.09	.884	-0.14	.810	0.06	.849	2.00	.045
PD+ γ_{02}	-0.61	.087	-0.10	.876	-0.59	.287	-0.01	.964	1.26	.203
Classroom mean of fall scores γ_{03}	-0.04	.207	-0.01	.807	0.16	.005	-0.02	.695	0.19	<.001
Global instructional quality γ_{04}	0.25	.260	0.71	.118	0.57	.154	0.19	.351	1.10	.112
Child level										
Fall score γ_{10}	0.74	<.001	0.74	<.001	0.61	<.001	0.64	<.001	0.55	<.001
Age γ_{20}	0.94	.007	0.01	.991	0.75	.131	0.78	.008	2.47	.016
Maternal education γ_{30}	0.64	<.001	0.23	.212	0.68	<.001	0.75	<.001	1.40	<.001
Gender γ_{40}	0.50	.068	0.89	.068	0.96	.015	0.97	<.001	2.37	.001
White γ_{50}	0.44	.238	0.32	.559	0.42	.355	0.18	.495	1.66	.063
Variance										
Intercept τ_{00}	3.03	<.001	6.72	.002	4.00	<.001	2.00	<.001	18.16	.002
Residual σ^2	25.923	<.001	82.65	<.001	30.99	<.001	18.36	<.001	199.57	<.001

Note. Results based on the full available sample and 40 imputed datasets. With one exception, all child-level variables were entered as fixed effects based on initial models indicating no significant random effects for these variables; the exception was for oral language, in which a random effect of fall language skill was included (estimate = 0.018, $p = .006$). None of the comparisons were statistically different after applying the linear step-up procedure to control familywise error rate. ICC = intraclass correlation from unconditional (empty) model.

Table 4

Multilevel results for PD and PD+ versus comparison on children's follow-up emergent literacy skills and kindergarten readiness

	Oral Language		Letter Recognition		Letter Sounds		Phonological Awareness		Print Concept Knowledge		Kindergarten Readiness	
	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*p*	estimate	*p*
Unconditional ICC	.25		.16		.23		.18		.20		.18	
Classroom level												
Intercept γ_{00}	50.02	<.001	42.56	<.001	16.28	<.001	12.67	<.001	125.48	<.001	21.58	<.001
PD γ_{01}	-0.40	.347	0.15	.833	-1.14	.133	-0.08	.838	-0.10	.933	-0.19	.688
PD+ γ_{02}	-0.17	.674	0.44	.532	0.27	.691	0.16	.657	1.42	.237	0.23	.624
Classroom mean of fall scores γ_{03}	-0.08	.015	-0.03	.367	0.02	.818	-0.02	.749	-0.06	.276		
Global instructional quality γ_{04}	0.14	.592	-0.01	.991	0.03	.959	0.09	.708	0.07	.935	0.33	.271
Child level												
Fall score γ_{10}[a]	0.62	<.001	0.39	<.001	0.36	<.001	0.58	<.001	0.47	<.001		
Age γ_{20}	0.66	.064	1.58	.009	1.77	.002	1.59	<.001	3.26	.001	-0.13	.745
Maternal education γ_{30}	0.33	.003	0.36	.044	0.35	.077	0.67	<.001	1.07	<.001	0.45	.001
Gender γ_{40}	0.07	.788	0.95	.052	1.00	.019	0.53	.043	1.02	.195	0.74	.031
White γ_{50}	0.90	.022	-0.18	.741	0.73	.214	-0.21	.510	1.32	.164	-0.57	.181
Variance												
Intercept τ_{00}	7.71	<.001	12.19	<.001	9.77	<.001	3.59	<.001	42.92	<.001	1.41	.081
Residual σ^2	22.19	<.001	84.10	<.001	34.67	<.001	23.11	<.001	202.59	<.001	14.48	<.001

Note. Results based on the full available sample and 40 imputed datasets. With one exception, all child-level variables were entered as fixed effects based on initial models indicating no significant random effects for these variables; the exception was for oral language, in which a random effect of fall language skill was included (estimate = 0.023, $p = .001$). None of the comparisons were statistically different after applying the linear step-up procedure to control familywise error rate. [a]The kindergarten readiness model also included four fall emergent literacy scores as covariates: oral language (coefficient = 0.243, $p < .001$), letter recognition (coefficient = 0.138, $p < .001$), phonological awareness (coefficient = 0.101, $p = .005$), and print concept knowledge (coefficient =

0.011, p = .370). ICC = intraclass correlation from unconditional (empty) model.

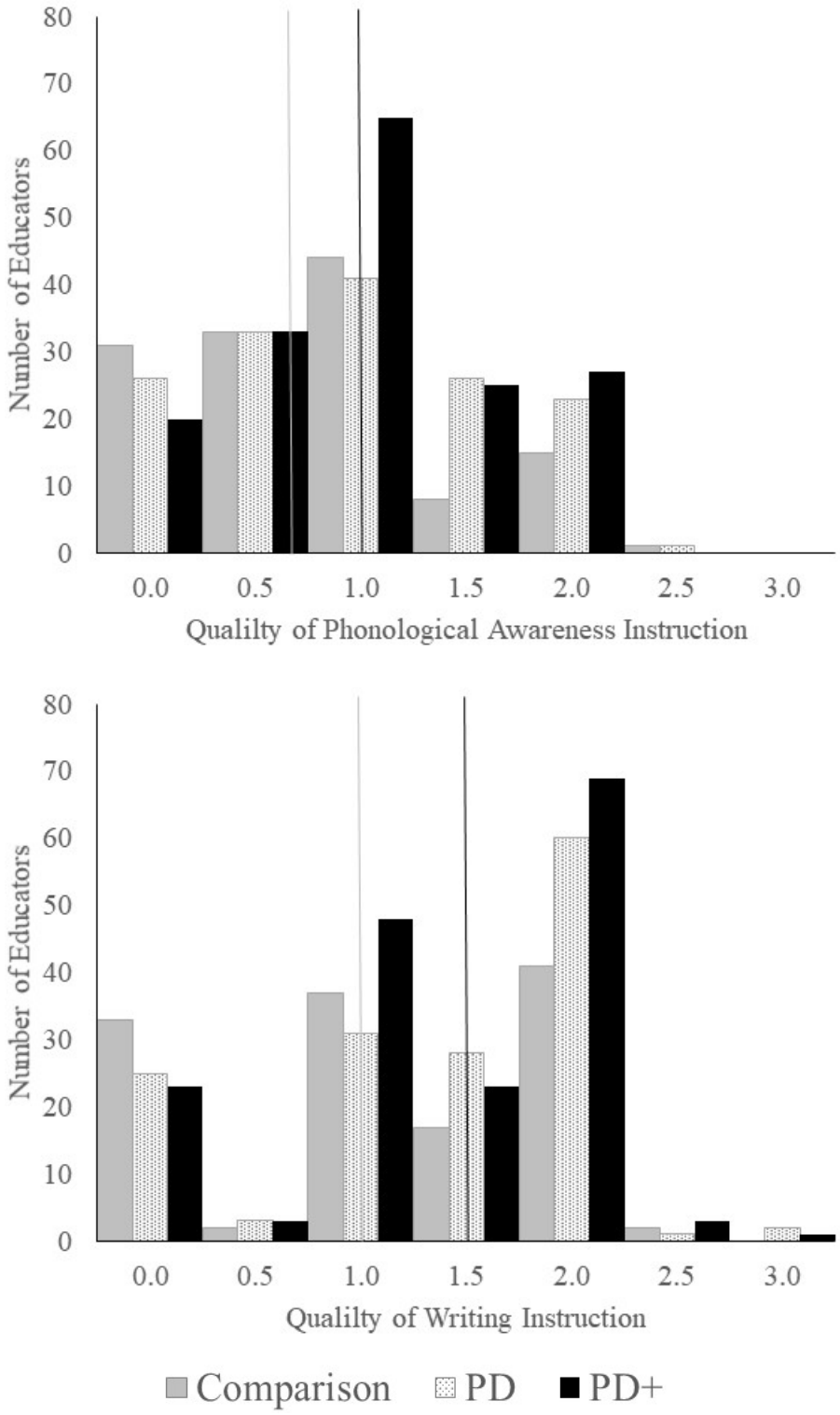

Figure 1. Histograms for quality of phonological awareness and writing instruction by condition. Gray vertical lines represent the median for comparison classrooms (0.67 and 1.00 for phonological awareness and writing, respectively), and black vertical lines represent the medians for PD and PD+ classrooms which were the same across conditions (1.00 and 1.50 for phonological awareness and writing, respectively).