Cost-Effectiveness of Early Childhood Interventions to Enhance Preschool: Evidence from a Randomized Experiment in Head Start Centers Enrolling Historically Underserved Populations

David S. Knight
Susan Landry
Tricia A. Zucker
Emily C. Merz
Cathy L. Guttentag
Heather B. Taylor

Abstract

We evaluate the cost-effectiveness of two early childhood interventions that use instructional coaching and parent coaching as levers for improvement. The study design allows us to compare the individual effects of each intervention as well as their combined effect on student outcomes. We find that teachers receiving instructional coaching improve their use of evidence-based instructional practices, while families receiving parent coaching show increases in numerous responsive parenting behaviors associated with positive child outcomes. Both interventions demonstrate positive impacts on students, but only parent coaching shows statistically significant effects across a range of student outcomes. Instructional coaching alone is substantially less costly and may therefore be the most cost-effective of the three treatment conditions; however, small sample sizes limit our ability to reach definitive conclusions. Policy simulations suggest that implementing these interventions could raise the overall cost-effectiveness of Head Start by at least 16 percent. © *2019 by the Association for Public Policy Analysis and Management.*

INTRODUCTION

Early childhood interventions such as preschool programs are one of the most effective means for improving school readiness and students' long-term outcomes, particularly for marginalized youth (Barnett, 2007; Campbell et al., 2014; Dynarski, Hyman, & Schanzenbach, 2013; Heckman et al., 2010). Many policymakers have called for the provision of universal prekindergarten (e.g., Obama, 2013). The U.S. Congress commissioned a national study of the federal Head Start preschool program as part of its reauthorization in 1998 to determine the effect on children's short- and long-term outcomes. One of the key findings of the Head Start Impact Study is that while the Head Start program provides benefits for enrollees, there is substantial variation in program quality (U.S. Department of Health and Human Services, 2010). Full-day service and frequent home visiting are two general characteristics linked to higher quality (Walters, 2015), although adding these practices can be expensive and expansion of funding for Head Start is limited (Knight, 2017; Mead, 2017). While many studies examine the

Journal of Policy Analysis and Management, Vol. 38, No. 4, 891–917 (2019)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
DOI:10.1002/pam.22145

effects of add-on preschool interventions, few rigorously analyze costs (Levin, 2001). Thus, a vital and overlooked role of research on preschool programs is to identify cost-effective means for improving the effectiveness of Head Start and other early childhood programs.

This paper compares the cost-effectiveness of interventions that combine instructional coaching for preschool teachers with family coaching for the parents of students enrolled in preschool. We analyze two widely implemented interventions, The Early Education Model (TEEM) and the Play and Learning Strategies (PALS) model. The design of our study allows us to test both the individual effects of these interventions and their combined impacts on a set of cognitive and non-cognitive student outcomes. The TEEM intervention involves trained coaches collaborating with Head Start teachers and directors to implement an evidence-based instructional program with developmentally appropriate lesson plans and teaching strategies. Similarly, the PALS intervention is aimed at improving responsive parenting techniques by matching parents with family coaches who provide instruction on specific interactive strategies as well as individualized coaching over the course of a school year. Henceforth, we refer to TEEM and PALS as instructional and parent coaching models, respectively, because of their emphasis on coaching as a mechanism for improvement.

We randomly assign the TEEM intervention to teachers across 77 Head Start centers over a three-year period. Within classrooms, students are randomly assigned to have their parents receive PALS. In total, we randomly assign 623 students to both the TEEM and PALS interventions, just one of these interventions, or the control group. Approximately 30 percent of students leave the study, leaving 434 in our final sample; however, attrition is roughly equal across treatment groups and summary statistics provide evidence that this attrition does not bias our impact estimates. We track a wide array of pre- and post-intervention outcomes for teachers, parents, and students. Finally, we conduct a rigorous cost analysis of each intervention using the ingredients method (Levin et al., 2017). Our study addresses the following research questions: What is the individual impact of the TEEM and PALS interventions and are there additional benefits for students who receive both interventions? What are the costs associated with each of the two interventions? Which of the three treatment conditions is the most cost-effective?

We find that both TEEM and PALS lead to positive and significant improvements across a range of teacher and parent outcomes. Both coaching models also demonstrate positive impacts on student outcomes, but the effects of TEEM and the combined intervention of TEEM plus PALS are not statistically significant (Landry et al., 2017). Although the parent coaching model (PALS) has slightly larger effects on student outcomes, parent coaching is over five times as costly as instructional coaching. Our results, therefore, suggest that instructional coaching alone (the TEEM-only condition) may be the most cost-effective of the three treatment conditions; however, small sample sizes limit our ability to draw definitive conclusions about which treatment condition is more cost-effective. Based on prior analyses of the overall cost-effectiveness of Head Start (Ludwig & Phillips, 2008; Puma et al., 2012), implementing parent and instructional coaching models in Head Start centers could increase the cost-effectiveness of Head Start by at least 16 percent.[1] Given our sample population, these findings generalize to Head Start centers serving

[1] The change in the cost-effectiveness of Head Start is calculated based on prior research of the cost-effectiveness of Head Start. Ludwig and Phillips (2008) find that Head Start has a total impact on student outcomes of 0.247 standard deviations (SD) with a total cost of $10,517 per student and an effectiveness-cost ratio of 0.0235 SD/$1,000. We find that implementing PALS increases the effect on student outcomes by 0.124 SD, compared to a Head Start control group, at a cost of $3,131 per student.

greater concentrations of Latinx and African American children relative to national averages.

The study's findings have important policy implications for enhancing the impact of Head Start. Instructional coaching programs are far less costly than home visit programs but may not always produce positive effects on students. Conversely, although one-on-one parent coaching models are expensive, their impact may warrant the cost. The study also reveals critical questions for future research. Would more intensive investment in teacher coaching improve Head Start teachers' instructional quality enough to positively impact students? Is the effect of home visit programs such as PALS large enough to justify the significant cost, relative to other potentially effective interventions for improving early childhood outcomes? In the balance of this paper, we review relevant background research and policy context, and then describe the experimental design of our study, the analytic approach and findings, and conclude with discussion and policy implications.

BACKGROUND ON INTERVENTIONS TO ENHANCE PRESCHOOL

Our primary research questions focus on the individual effects of parent and teacher coaching interventions on children's school readiness, their combined effects, and which of three treatments (parent coaching only, teacher coaching only, or both) is most cost-effective at increasing student measures of school readiness. Below, we review literature that addresses these questions. We first provide background on how school readiness is defined in prior studies. We then discuss extant research assessing the impact of parent and teacher coaching models, including TEEM and PALS, on school readiness. We offer a rationale for implementing parent and teacher coaching interventions simultaneously and present our hypotheses of their individual and synergistic effects on students. Finally, we describe early childhood research that draws on cost analysis and describe why the dearth of cost-effectiveness analyses in particular represents an important gap in the literature.

Defining School Readiness

Child development literature defines indicators of school readiness that predict students' success in later grades (e.g., Brown, 2013; Lee et al., 2014). The broader domains of school readiness include (a) cognitive skills, (b) social and behavioral skills, (c) self-regulatory processes, and (d) executive function (Garon, Bryson, & Smith, 2008; Mashburn & Pianta, 2006). Cognitive skills include oral language and early literacy and numeracy skills such as letter and number knowledge and phonological awareness (Noble, Norman, & Farah, 2005), while social and behavioral skills include cooperation, social engagement, and comfort with new situations. Self-regulatory and executive function skills involve the ability to focus attention on goal-directed activity, resist interference from external stimuli, respond effectively to situations that are motivationally significant, and regulate emotions (Garon, Bryson, & Smith, 2008). Researchers have developed and tested interventions that provide structured learning opportunities for teachers or parents to support children's school readiness (e.g., Burns, Donovan, & Bowman, 2000). We describe research on these interventions next.

In other words, Head Start with PALS would have an overall effect of 0.371 SD (0.247 0.124), an overall cost of $13,648 ($10,517 $3,131), and an effectiveness-cost ratio of 0.027 SD/$1,000 per pupil, which represents a 15.6 percent increase. Results for TEEM imply an overall effectiveness-cost ratio for Head Start of 0.034 SD/$1,000 per pupil, a 30.7 percent increase.

Research on the Impact of Parent and Teacher Coaching

Two broad characteristics linked to early childhood program effectiveness are high-quality instruction and interventions that involve home visits (e.g., Barnett, 2011; Heckman & Kautz, 2013; Wong et al., 2008). Sweet and Appelbaum (2004) review 60 home visit programs and find that while most significantly improve student outcomes, not all programs are effective at increasing school readiness. Programs that specifically emphasize increasing parent responsiveness and support for school readiness are linked to longer-term student achievement (e.g., Van Zeijl et al., 2006). Bierman et al. (2015) find in a randomized trial that home visits that provide parents with evidence-based learning games and guided pretend play to use with their children improve child literacy skills and academic performance prior to and during the kindergarten year. Two other studies randomly assign families to receive either home visits involving training on responsive parenting techniques or paper materials with similar information (Bakermans-Kranenburg et al., 2003; Van Zeijl et al., 2006). Parents receiving home visits increase their use of responsive parenting techniques and their children show gains in independent problem-solving, language, social and emotional skills, and behavioral development. PALS is a responsive parenting intervention, evaluated in the current study, that involves ongoing parent coaching on specific evidence-based parenting strategies. Parenting strategies include those relevant to school readiness such as stimulating language development and cognitive problem-solving, reading books in an interactive, engaging manner, and maintaining children's attentional focus. Several previous random-assignment efficacy studies find significant positive effects of PALS on parental responsiveness behaviors as well as children's language skills, cooperation, and social engagement (Guttentag et al., 2014; Landry et al., 2008; Landry et al., 2011a).

Studies show that early childhood teachers also play an important role in contributing to students' cognitive and social-emotional development (Burns, Donovan, & Bowman, 2000). In summarizing experimental research on preschool programs, Barnett (2011) notes that highly effective programs—those that exceed the impacts of Head Start—hire college-educated teachers and pay salaries equivalent to public school teachers. Although some studies find no relation between teacher credentials and student outcomes (Early et al., 2006; Walters, 2015), teaching quality is likely a strong predictor of preschool effectiveness. Studies also link teacher training on specific pedagogical strategies or curricular programs to greater development of academic skills and social, behavioral, and self-regulatory processes (e.g., Chiang, Clark, & McConnell, 2017; Raver et al., 2011). Students of teachers who are randomly assigned to training on the curricula for Promoting Alternative Thinking Strategies (PATHS), for example, show higher socio-emotional skills and are rated by parents and teachers as more socially competent compared to peers (Domitrovich, Cortes, & Greenberg, 2007). TEEM is a similar teacher professional development intervention, evaluated in the current study, that involves student progress monitoring and instructional coaching to facilitate preschool teachers' use of a specialized curriculum. Two previous large-scale experimental studies find that TEEM increases teachers' use of evidence-based, responsive pedagogical practices, which improves children's language and literacy skills (Landry et al., 2009, 2011b). While our prior research finds positive effects for both PALS and TEEM, we have not studied the effects of these interventions when implemented concurrently.

Theoretical Basis for Providing Complementary Interventions

Despite substantial research on the effects of parent-based interventions and teacher professional development in early childhood settings, including both TEEM and PALS, fewer studies examine the effects of "parent-plus-teacher" interventions

(Bierman et al., 2017; Lonigan & Whitehurst, 1998; Whitehurst et al., 1994). Yet there is a strong theoretical rationale for interventions that emphasize consistency across the home and school environments. Theories of person-environment fit (Bronfenbrenner & Morris, 2006) suggest that providing a consistent atmosphere focused on similar child behaviors may help children better develop cognitive and non-cognitive skills. If home visits support particular parenting strategies—such as responsive parenting—but those lessons are not reinforced in preschool settings, children may receive mixed messages about desired behaviors. Conversely, children may benefit more from learning experiences in preschool if caregivers emphasize similar lessons at home. The interventions evaluated in this study are intended to complement each other because each provides training to teachers or parents that supports the same sets of child skills associated with school readiness.

We implement the TEEM and PALS interventions on a parallel timeline. Both interventions target cognitive (literacy and language skills) and non-cognitive (school liking, engagement, and social-emotional and executive functioning) child outcomes. We hypothesize that because lessons students learn in their preschool classroom are reinforced at home, implementing these interventions in unison may provide either additive or synergistic effects. Synergistic effects may arise if the impact of receiving both interventions is greater than the sum of receiving each individual intervention. Alternatively, the two interventions could simply provide additive effects when implemented together, or they could be redundant in that providing both interventions create the same impact as providing one or the other. Given differences in resource requirements, any one of the three treatments, parent coaching only, teacher coaching only, or parent and teacher coaching, could be the most cost-effective. The limited prior research on parent-plus-teacher interventions finds mixed (Lonigan & Whitehurst, 1998; Whitehurst et al., 1994) or positive (Bierman et al., 2017) results, but none of these studies consider costs.

Discerning among the alternate hypotheses outlined above is important for programmatic decisionmaking and policy reform. Preschool providers are faced with important resource allocation decisions within their preschool centers. These providers would benefit from knowing whether parent and teacher coaching models have redundant, additive, or synergistic effects on students. Findings may also inform state and federal regulations on resource allocation in publicly funded prekindergarten or Head Start centers. States that are considering expanding their funding for preschool may need to consider whether teacher or parent coaching is more cost-effective, in addition to which model is more effective. Policymakers may be interested in supporting both models; however, if the effects of TEEM and PALS are redundant, then Head Start centers and other preschools should avoid using both the teacher and parent coaching models and instead implement whichever is more cost-effective. If the interventions have additive effects, then implementing both programs may be beneficial. Finally, if the programs have synergistic effects, then centers implementing the TEEM coaching program could be encouraged to implement PALS since doing so would increase the impact of TEEM. That said, even under the presence of synergistic effects, implementing both programs may not be the most cost-effective approach if one program requires significantly more resources than the other. This study alone does not provide definitive answers to these questions but offers important evidence about effective resource allocation decisions in early childhood settings.

Cost-Benefit and Cost-Effectiveness Analysis of Early Childhood Education

While few studies of preschool employ cost-effectiveness analysis, many involve cost-benefit analyses, which demonstrate the positive social returns to early childhood interventions (Barnett, 1985, 2007; Heckman et al., 2010). The difference

between these two methodologies is important. A cost-benefit analysis converts all inputs and outputs of a policy or program into monetary terms to answer the question of whether the social benefits exceed the social cost, as measured in dollar values (Aos & Pennucci, 2013; Boardman et al., 2018). A particular program can be assessed with respect to its social return on investment. Cost-effectiveness analysis differs in that outcome measures are not converted to dollar figures and instead are reported in measures of student achievement or some other relevant outcome (Levin et al., 2012). In contrast to cost-benefit analyses, cost-effectiveness analyses are inherently comparative because the cost of raising achievement, for example, must be compared to some relative alternative in order to have meaning (Levin et al., 2017).

One of the strengths of cost-effectiveness analysis is that it uses standardized methods that, to some extent, allow researchers to draw comparisons across studies (e.g., Harris, 2009; Levin, 2001). Cost-effectiveness studies often use standard deviations (SD) of student achievement, or effect sizes, to measure outcomes. While Cohen (1988) defines small, medium, and large effect sizes as 0.2, 0.5, and 0.8, scholars question this classification as overly simplistic and unrealistic (Kraft, 2018; Lee, Finn, & Lui, 2012). Moreover, these benchmarks ignore costs. Reviewing a broad set of studies of educational interventions that include estimated effect sizes and information about costs, Harris (2009) suggests that any intervention that increases test scores by 0.025 SD per $1,000 per student per year could be considered large. This ratio is informative because it allows researchers to make comparisons of the cost-effectiveness of interventions evaluated across studies. Analyses of class size reduction in lower elementary grades suggest that each $1,000 spent on class size reduction increases math and reading achievement by between 0.026 and 0.086 SD (Harris, 2009; Krueger, 2003; Levin, Glass, & Meister, 1987). While most cost studies of Head Start utilize cost-benefit analysis, Ludwig and Phillips (2008) use first-year findings of the National Head Start Impact Study (Puma et al., 2005) to estimate the cost-effectiveness of Head Start. They find that Head Start has an average effect size for cognitive and non-cognitive outcomes of 0.247 (based on Table 1 of Ludwig & Phillips, 2008, and Exhibit 1 of Puma et al., 2005) and a cost of $10,517 per child (in 2018 dollars, although no formal cost analysis is conducted). The authors conclude that the program raises achievement by approximately 0.023 SD per $1,000 per child. Therefore, interventions that both increase the impact of Head Start and make the program more cost-effective should have effectiveness-cost ratios larger than 0.023 SD per $1,000 per pupil. Despite the prevalence of cost-benefit studies demonstrating that preschool is a sound social investment overall, no other studies that we know of have used cost-effectiveness analysis to evaluate interventions designed to enhance the efficiency of Head Start.

INTERVENTIONS RESEARCH DESIGN AND ANALYTIC APPROACH

In this section, we first discuss the interventions being studied, TEEM and PALS. We then describe the experimental design, our data collection procedures, and the analytic approach for estimating causal effects and assessing cost-effectiveness.

The Early Education Model and the Play and Learning Strategies Interventions

TEEM consists of a package of teacher professional development resources including: (a) a two-day initial training; (b) instructional coaching both in class and through web-based professional development courses; (c) student progress-monitoring; and (d) instructional resources. We refer to TEEM as an instructional coaching model, since coaching represents the largest investment from a resource perspective. The purpose of the TEEM intervention is to increase the use of

instructional techniques that promote students' cognitive growth and social-emotional skills by providing rich language input, maintaining children's attentional focus, balancing teacher- and child-directed activities, and providing contingent responses. A two-day initial training introduces teachers to the components of the TEEM framework and the topics covered in coursework and coaching sessions (see Table A1).[2] The coursework takes place over approximately 20 bi-weekly two-hour sessions, led by an instructional coach. Coursework involves a web-based platform that facilitates group discussion and role-play of evidence-based instructional techniques. As part of the coursework, teachers are trained to administer progress monitoring of students three times per year and to use those results to guide individualized instruction. Teachers are provided with laptops to ensure access to the web-based courses, complete "homework" activities, and administer progress-monitoring measures.

The PALS intervention takes place over the course of a school year, through 16 to 20 one-on-one bi-weekly coaching sessions, to help parents implement research-based responsive parenting techniques. The PALS curriculum materials include a manual and a set of training videos that coaches use to help parents learn responsiveness behaviors. During each PALS session, coaches video record the parent implementing the responsive behaviors in an activity with their child (e.g., book reading, play, or mealtime) and the parent then views the video with the coach and critiques their own behavior and how their child responded. Responsive parenting behaviors involve following the child's lead in play, attending to and understanding communicative signals, responding promptly and contingently, using scaffolding strategies to support language development, maintaining rather than redirecting the child's focus of attention, and using positive behavior management strategies during play and learning activities.

Research Design

The study involves 77 Head Start classrooms in the Houston and Austin metropolitan areas over three years, from school year 2009/2010 to 2011/2012. We contacted directors of six major Head Start agencies and all agreed to participate. Parents of children in all classrooms are invited to participate and sign informed consents. On average, more than 70 percent of parents consent to participate in the study, resulting in 857 consenting families. We then randomly assign half of all classrooms to receive the TEEM intervention and assign the other half to a business-as-usual control condition. We cluster randomization by city to ensure an equal number of treatment classrooms are located in Houston and Austin. Of the children with parental consent, six to eight children per classroom are randomly selected for participation (623 total). Half of these students (314) are randomly assigned to have their parents receive PALS and the other half (309) are assigned to the control condition. This 2×2 design creates four student treatment groups: (a) TEEM and PALS; (b) TEEM, no PALS; (c) no TEEM, PALS, (d) no TEEM and no PALS (control). The interventions take place over one school year and a new group of teachers, parents, and students is recruited each year.

Teacher and student demographics for the final sample after attrition are shown in Table 1. None of the differences in demographic characteristics across treatment groups are statistically significant. Students are primarily Latinx (70 percent) and African American (29 percent) and a small proportion are White or Asian. Approximately 39 percent of students in each condition speak mostly Spanish at home,

[2] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Table 1 Demographic information for teachers and students by treatment group.

	Teacher/classroom			Students				
	TEEM	No TEEM	Total	TEEM and PALS	TEEM, no PALS	PALS, no TEEM	Control	Total
Sample size	39	38	77	102	112	109	111	434
Age, gender, and race/ethnicity								
Average age in years (SD)	–	–	–	4.42 (0.46)	4.38 (0.54)	4.34 (0.50)	4.40 (0.50)	4.38 (0.50)
Female	94.9%	94.6%	94.7%	47.6%	51.9%	52.4%	48.1%	48.6%
African Amer.	48.7%	64.9%	56.6%	24.8%	31.8%	30.3%	29.1%	29.1%
Latinx	33.3%	21.6%	27.6%	74.3%	64.6%	68.8%	70.9%	69.5%
Caucasian/White	12.8%	10.8%	11.8%	1.0%	3.6%	0.0%	0.0%	1.2%
Asian	2.6%	2.7%	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%
Experience and highest level of education (teachers only)								
Avg. years of teach. exp. (SD)	11.29 (9.03)	10.65 (9.11)	10.97 (9.07)	–	–	–	–	–
High sch./GED	10.3%	0.0%	5.3%	–	–	–	–	–
CDA credential	28.2%	18.9%	23.7%	–	–	–	–	–
Associate's deg.	12.8%	16.2%	14.5%	–	–	–	–	–
Bachelor's deg.	41.0%	46.0%	43.4%	–	–	–	–	–
Some grad. sch.	7.7%	19.0%	13.2%	–	–	–	–	–
Family income (students only)								
FRL status	–	–	–	96.0%	96.4%	96.3%	98.2%	96.1%
Language spoken in the classroom or home								
Mostly English	66.7%	75.7%	71.1%	39.2%	49.1%	40.7%	46.9%	44.0%
Eng. and Sp.	28.2%	16.2%	22.4%	20.6%	15.2%	16.7%	14.4%	16.6%
Mostly Spanish	5.1%	8.1%	6.6%	40.2%	35.7%	42.6%	38.7%	39.2%

Notes: All teachers, parents, and students in the study identified as one of the four race/ethnicities listed. TEEM stands for The Early Education Model, PALS stands for Play and Learning Strategies, FRL stands for free or reduced-price lunch, and SD stands for standard deviation. The control group did not receive the TEEM or the PALS intervention. Average parent characteristics are also similar across the PALS and no PALS conditions.

while all other students speak a mix of English and Spanish (17 percent) or mostly English (44 percent). Most students are between three to five years old, with an average of 4.38. Parents report that 96 percent of children qualified for free lunch. The majority of teachers are African American and Latinx and teachers have 11 years of teaching experience, on average. Approximately one-fifth of teachers deliver some of their lessons in Spanish and 7 percent use mostly Spanish.

The study includes 21 family coaches and two teacher coaches. We take significant care to ensure high fidelity of implementation. Both the TEEM and PALS coaching components are highly scripted and could be replicated by other coaches who complete the requisite training. We measure fidelity by tracking the extent to which teachers and parents participate in the various components of TEEM and PALS. Results of our fidelity measures are reported in Table A2. Senior research staff also supervise coaches during weekly group meetings and monthly home visits, activities that are typical when TEEM and PALS are implemented in non-research settings.

Finally, we closely monitor student and teacher attrition. We initially recruit and administer pretests for 623 families and students and 434 are included in the final analysis (30 percent attrition rate). Parents or students leave the study because parents sought other programs such as public school prekindergarten or because they moved residences. In Table A3, we report summary statistics for students who leave the study after initial recruitment and those who remain in the study. We find no meaningful or statistically significant differences in attrition across treatment conditions or in pretest scores or demographic characteristics for those who leave and those who remain in the study.

There are five TEEM teachers and four non-TEEM teachers who leave the study, a total of nine of the 79 teachers. We replace six of these teachers in time for pretests and a seventh early in the school year (we exclude this teacher in analyses of teacher effects). Several pieces of evidence suggest the two teacher groups remain equal in expectation after attrition and replacement. First, in follow-up interviews, exiting teachers report leaving for reasons unrelated to the intervention or their expected outcomes, including classroom dissolution, transfer to a different classroom, or leaving the school. Second, pretest scores of the teacher observational ratings show no meaningful difference between those who leave the study, replacement teachers, and those who remain in the study. Teachers who leave the study also had similar experience and educational levels as the replacement teachers and also as those who remain in the study, as shown in Table A4. Although we are unable to rule out the possibility that a teacher's decision to leave the study is related to their research assignment, the fact that roughly the same number leave from each group and that leavers and stayers have similar characteristics provides some evidence that teacher attrition is unrelated to research assignment. Finally, we recruit replacement teachers from the same Head Start center from which the previous teacher has left. We use this same approach regardless of whether a teacher has been in the TEEM or the no-TEEM condition.

Methods for Assessing the Effects of TEEM and PALS

Teacher and Parent Outcomes

To track instructional change associated with TEEM, we conduct classroom observations at the beginning, middle, and end of each school year for treatment and control teachers, using the Teacher Behavior Rating Scale (TBRS; Landry et al., 2009). Classroom observers receive certification in coding of the TBRS, which includes demonstrating mastery of codes and achieving 80 percent reliability with coded videotaped lessons and side-by-side live coding with master coders. We also confirm reliability of classroom observations by having two raters for 15 percent of all classroom observations. The TBRS includes a total of 61 items that measure quantity and quality of 11 subscale areas such as Classroom Community, Oral Language, Book Reading Behaviors, and others (see Table A5).[3] Each subscale is measured on a scale from zero to four. We estimate separate models predicting mid-year and end-of-year assessment scores, controlling for pretests and relevant teacher characteristics. The coefficient for the TEEM indicator variable provides the causal impact of TEEM professional development on teacher behaviors.

We assess parent outcomes for PALS through observer ratings of videotaped parent-child interactions during scheduled free play and book reading sessions (with

[3] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

toys and books provided by family coaches). Free play sessions involve parents interacting with their children with toys, while book reading sessions involve parents reading books to their children. Parents are observed before and after implementation of the PALS intervention in their home by trained observers who are blind to the intervention status. Observers spend three weeks practicing coding, with the requirement that their ratings are 80 percent in agreement with a master coder. We hold monthly meetings with observers to review and code videotapes as a group and include a second rater for approximately 15 percent of all parent observations. Observation instruments for book reading and free play sessions are developed by the authors in prior work (Landry et al., 2011a; Landry et al., 2009). We find these instruments to be reliable and valid measures in evaluations of other interventions including "CATCH," a child health program, as well as alternate versions of PALS designed for parents of younger children or children with special needs. All parent observation ratings are factor-analyzed to assess overlap in measures and, in some cases, a single factor is identified from multiple constructs. Additional information on parent observation outcomes is included in Table A6. We model the impact of PALS on parent outcomes using simple OLS regression, predicting end-of-year scores based on pretests and treatment status.

Student Outcomes

We track 17 separate student outcome measures at the beginning and end of each school year. Student outcomes include four domains: (a) cognitive skills; (b) social and behavioral skills; (c) self-regulatory processes; and (d) executive function. These outcomes align with the intended outcomes of the interventions and are based on extant school readiness research (Brownell, 2001; Lonigan et al., 2007; Zimmerman et al., 2002). We assess outcomes using four different techniques:

> Observation ratings of student behavior conducted by researchers during child-parent book reading sessions and free play sessions (six outcomes: book reading engagement, language use, shared enjoyment, enthusiasm/initiative, cooperation, and social engagement);
> Student "tasks" (three outcomes that measure self-regulation and executive functioning);
> Teacher and parent surveys (five outcomes that assess social-emotional functioning, school liking, and school avoidance); and
> Standardized, norm-referenced tests (three outcomes that assess language and literacy skills).

Initially, outcomes based on standardized, norm-referenced tests included a total of six outcomes: three subsets of the Test of Preschool Early Literacy (TOPEL), two subscales of the Preschool Language Scale, Fourth Edition (PLS-4), and the Expressive One-Word Picture Vocabulary Test (EOWPVT). Due to the high correlations among the language measures, we form a composite "language skills" outcome consisting of the EOWPVT, both PLS subscales, and the Definitional Vocabulary subscale of the TOPEL. We assess intraclass correlations for all observation measures and Cronbach's alpha for all survey measures. These assessments demonstrate that our observation and survey measures are reliable and internally consistent, respectively, according to conventions established in the literature (Rust & Golombok, 2009).[4] In sum, we measure a total of 17 cognitive and non-cognitive

[4] We provide a more complete description of all student outcome measures, including our assessments of internal consistency and reliability in the Appendix text and in Table A6. Table A7 provides descriptive

student outcomes using four different data collection procedures. We report all outcomes in both raw coefficients and Cohen's *d* effect sizes.

Modeling Student Outcomes and Interpreting Coefficients

We estimate the causal impacts of each intervention on student outcomes using separate hierarchical linear models predicting each outcome based on pre-score and treatment status, including the interaction of the two treatments. We use the notation described in Raudenbush and Bryk (2002), in which the first subscript refers to the level 1 variables and the second subscript refers to level 2 variables. The treatment indicator for TEEM and its interaction with the PALS treatment indicator are classroom-level variables, while the treatment indicator for PALS is a student-level variable, indexed by *j* and *i*, respectively:

$$Y_{ij} = \gamma_{00} \quad \gamma_{10} PRESCORE_{ij} \quad \gamma_{20} PALS_{ij} \quad \gamma_{01} TEEM_j \quad \gamma_{21} TEEM_j * \text{PALS}_{ij} \quad \mu_{0j} \quad r_{ij}$$

Positive coefficients for the main effects of the TEEM and PALS treatment variables (γ_{01} and γ_{20}, respectively) indicate that students who received only the teacher or family coaching intervention outperformed the control group. Positive coefficients on the interaction term, γ_{21}, indicate synergistic effects of receiving both interventions.

Synergistic effects exist when students who receive both interventions have higher predicted outcomes than would be expected from summing individual effects of each intervention. If γ_{21} is zero, we assume that combining the two programs does not provide additional benefits over and above the sum of the two, but the two programs are not replicative of each other (the programs are additive in that each intervention individually contributes to measured student outcomes). However, a negative γ_{21} implies that the two interventions are redundant and providing both leads to lower measured outcomes than would be expected by summing the effects of each. We allow residuals to vary at the classroom and student level, denoted by μ_{0j} and r_{ij}, respectively, and set alpha equal to 0.100. Given the multiple non-independent outcomes tested, we adjust all significance tests using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995; Hochberg, 1988).[5] Before presenting our results, we first review our methods for assessing costs.

Methods for Assessing the Cost of TEEM and PALS

We measure the total and per-pupil yearly cost of implementation using the ingredients method (Levin & Belfield, 2015; Levin et al., 2017). We collect data about personnel staff allocations and non-personnel resources used in each intervention

statistics for student pre- and post-intervention outcomes. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

[5] The Benjamini-Hochberg procedure reduces the rate of false-positive findings back to the level of an individual significance test. The method involves rank ordering the point estimates for each intervention by smallest to largest *p*-value, from m=1 to k (where k is the total number of significance tests and m is each coefficient's rank order of *p*-value, from smallest to largest), and rejecting the null hypothesis when the *p*-value is less than alpha * k/m. For example, for the first hypothesis test in a set of 17 hypothesis tests (i.e., 17 student outcomes), adjusted alpha is 0.1 * 1/17 = 0.006. Adjusted alpha for the second hypothesis test is 0.1 * 2/17 = 0.012. We implement the Benjamini-Hochberg method separately for teacher, parent, and student outcomes, since each domain of outcomes is based on independently collected data and separate analytic frameworks.

through interviews with developers and research participants. Interviews are followed with e-mail correspondence to ensure that our estimates of resource use are accurate. Next, we assign dollar values for each resource based on their current market price, the actual salary for that employee (for coach trainers and supervisors), or the prevailing salary for an individual with similar qualifications. For physical resources that last multiple years such as cameras and DVD players, we annualize their value over their lifetime. For start-up professional development that coaches draw upon for multiple years, we discount costs to present value using a 3 percent discount rate. Additional information on the cost analysis methods is available in the Appendix.

Methods for Comparing Cost-Effectiveness

We use cost-effectiveness ratios to determine which intervention is more cost-effective. Cost-effectiveness ratios are calculated as the yearly per-student cost divided by the SD of effect size. For ease of interpretation, we also report effectiveness-cost ratios, which measure the total effect size for each $1,000 per student per year. The overall effect size associated with each treatment condition is the average effect size across all outcome measures. The per-pupil annual cost of TEEM and PALS is the average across sites. Because the TEEM and PALS interventions do not involve any sharing of resources, the annual per-pupil cost of implementing both interventions is simply the sum of the costs for each intervention. Although our initial hypotheses suggest that students will benefit from TEEM and PALS, and that students receiving both interventions will experience synergistic effects, which of the three treatment conditions is most cost-effective is unclear, *a priori*, given uncertainty in the expected magnitude of effects and expected cost.

RESULTS

Findings are presented in Tables 2 through 5. We discuss intervention effects, costs, and cost-effectiveness in the sections below.

Assessing the Effects of TEEM and PALS

Effects of TEEM

The TEEM intervention leads to observable instructional changes for teachers but has no statistically significant effects on student outcomes. Effects of TEEM on teachers' instructional behavior are shown in panel A of Table 2. Although treatment teachers outperform control group teachers on most TBRS subscales (and on the overall scale), TEEM teachers do not continue to show greater gains on the TBRS throughout the school year. Treatment teachers increase their overall score by 0.314 points more than control group teachers from pretest to midyear assessment, after adjusting for teacher characteristics (shown in row 1, column 1, of Table 2). However, from pretest to end-of-year assessment, treatment teachers increase their overall score by 0.241 points more than control group teachers, implying that the control group teachers make up some of the gap during the second half of the school year (row 1, column 3). These overall results generally hold for the TBRS subscales. Treatment teachers outperform control group teachers on almost all subscales from pretest to midyear assessment, but only maintain statistically significantly higher scores on three TBRS subscales (Learning Centers, Lesson Plans, and Print and Letter Knowledge) from pretest to end-of-year assessments.

Table 2 Regression coefficients and standardized effect sizes (Cohen's *d*) for teacher and parent outcomes associated with the TEEM intervention (panel A) and the PALS intervention (panel B).

Panel A: Observation scores on the TBRS					*Panel B: Observation scores of parent activities*				
	Timing of classroom observation					Activity observed			
	Mid-year		End-of-year			Shared book reading		Free play	
	b (SE)	ES	*b (SE)*	ES		*b (SE)*	ES	*b (SE)*	ES
Total	0.314 (0.080)	0.93**	0.241 (0.079)	0.71*	Total	–	0.41**	–	0.19***
TBRS subcategories					*Parent observation score subcategories*				
Book read. behavior	0.419 (0.118)	0.90**	0.192 (0.118)	0.41	Prompts	0.300 (0.136)	0.24*	3.887 (0.856)	0.44***
Learning centers	0.358 (0.159)	0.57*	0.427 (0.158)	0.68*	Praise	0.336 (0.160)	0.25*	1.318 (0.142)	0.53***
Classroom community	0.267 (0.128)	0.45*	0.226 (0.128)	0.38	Verbal scaffolding	0.586 (0.216)	0.32**	0.140 (0.253)	0.18*
Teacher sensitivity	0.244 (0.097)	0.68*	0.159 (0.097)	0.45	Text dur. time	23.886 (4.461)	0.59***		–
Lesson plans	0.382 (0.191)	0.47	0.382 (0.190)	0.47	Tracking print	0.354 (0.102)	0.38**		–
Math concepts	0.412 (0.154)	0.67*	0.196 (0.153)	0.32	Provision of labels		–	3.381 (1.000)	0.32**
Oral language	0.294 (0.125)	0.66*	0.209 (0.125)	0.47	Maintain attention		–	–0.921 (0.503)	–0.19
Print and letter knowl.	0.226 (0.122)	0.48	0.330 (0.122)	0.70*	Redirect focus of att.		–	–0.722 (0.187)	0.41***
Written expression	0.204 (0.129)	0.38	0.201 (0.129)	0.38	Resp. lang. & beh. sup.	–	0.44***	–	0.48***
					Lang. build. strat.	0.900 (0.249)	0.55**		–
					Negativity	0.928 (0.300)	0.56**	0.803 (0.267)	0.49**

Notes: All categories measuring an "undesirable" outcome (i.e., text duration time, maintain attention, redirect focus of attention, and negativity) are reverse coded so that positive numbers indicate a beneficial effect. For PALS outcomes observed during shared book reading, "responsive language and behavior support" is a composite score composed of the following scales: enthusiasm and engagement, responsiveness/flexibility, and language comprehension techniques. For PALS outcomes observed during free play, "responsive language and behavior support" is a factor score composed of the following scales: responsiveness/flexibility, positive affect, warmth, and verbal scaffolding. TEEM effects are based on a sample of 231 teacher-observations and PALS effects are based on 868 parent-observations for each activity. p<.1, *p<.05, **p<.01, ***p<.001, corrected for multiple comparisons (Benjamini & Hochberg, 1995).

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

TEEM teachers demonstrate effect sizes of 0.93 and 0.71 for the midyear and end-of-year assessments of the overall TBRS score, respectively.[6] Despite the modest gains on the TBRS scales for treatment teachers described above, when converted to effect sizes, the gains are relatively large compared to some studies of teacher professional development. For example, using other assessment tools to monitor teachers' instruction, Garet et al. (2008) and Garet et al. (2011) find effect sizes associated with teacher coaching of 0.53 and 0.48, respectively. That said, the effect size for our end-of-year assessment, 0.71, is not as large as we find in prior evaluations of TEEM. In a large experimental study that tracks teachers' instruction using the TBRS, we find effect sizes associated with TEEM of 0.84 after one year of receiving TEEM professional development and 1.71 after the second year (Landry et al., 2011b). In an earlier large-scale randomized control trial, we find an effect size for TEEM professional development of 1.11 on the overall TBRS score, with effect sizes on subscales ranging from 0.43 to 1.46 (Landry et al., 2009). In the current study, TEEM teachers did not surpass the raw score of 3.0 on most TBRS subscales. Thus, despite making statistically significant gains over control group teachers that also resulted in relatively large effect sizes, TEEM teachers do not show as much growth on the TBRS as we hypothesized.

The effects of the TEEM intervention on student outcomes are reported in the first row of each panel of Table 3. All outcomes are relative to the business-as-usual condition (no TEEM, no PALS). Of the 17 outcomes assessed, 13 are positive and eight have effect sizes of 0.10 or greater, but none are statistically significant when correcting for multiple hypothesis tests.[7] In short, the TEEM intervention leads to moderate changes in instruction, but students assigned to TEEM classrooms do not experience consistently greater gains in outcomes compared to the control group. The small sample size, especially for teachers, may limit our ability to detect nonzero impacts as statistically significant. In our discussion, we provide additional plausible explanations for the presence of positive teacher impacts and the general lack of positive and significant student outcomes.

Effects of PALS

Effects of the family coaching intervention for parents are shown in panel B of Table 2. We find that PALS positively impacts all of our observed parenting measures. The first raw coefficient, labeled Prompts, suggests that PALS parents increase the number of times they prompt their child to say or do something during book reading sessions by 0.30 instances more than parents not receiving PALS. Parents receiving family coaching through PALS increase their use of prompts by roughly a quarter of an SD over the control group parents (effect size of 0.24). The greatest impacts are seen on text duration time—a measure of the amount of time parents spent during book reading sessions in conversation beyond just reading the text (e.g., providing verbal scaffolding). The coefficient of 23.9 suggests that parents in the PALS condition decrease their time spent just reading text to their child during the 10-minute book reading sessions by about 24 seconds more than

[6] Pretest to midyear effect sizes on TBRS subscales range from 0.38 to 0.90 and pretest to end-of-year effect sizes range from 0.32 to 0.70. The overall effect sizes are larger than the effect sizes of individual subscales because the pooled SD of teachers' average TBRS prescores is smaller than TBRS subscales for both midyear and end-of-year assessments.

[7] Table A8 shows uncorrected results for interested readers. Four outcomes were positive and independently statistically significant prior to the Benjamini-Hochberg procedure. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Table 3 Effects of the TEEM and PALS interventions on student outcomes (beta coefficients, standard errors, and Cohen's *d*).

	Panel A: Student outcomes based on observational ratings of child behaviors											
	Book reading sessions								Free play sessions			
	Book reading engagement		Language use		Shared enjoyment		Enthusiasm/ initiative		Cooperation		Social engagement	
TEEM	0.206	0.18	0.248	0.20	0.265	0.21	0.717	0.23	0.114	0.07	0.060	0.04
	(0.129)		(0.157)		(0.128)		(0.361)		(0.205)		(0.205)	
PALS	0.323	0.27	0.350	0.28	0.346	0.28	1.009	0.32*	0.190	0.12	0.130	0.08
	(0.129)		(0.145)		(0.122)		(0.356)		(0.207)		(0.206)	
TEEM × PALS interaction	0.630	0.54	0.298	0.24	0.398	0.32	1.322	0.42	0.294	0.18	0.039	0.02
	(0.258)		(0.289)		(0.241)		(0.706)		(0.411)		(0.410)	
	Panel B: Student outcomes based on standardized, norm-referenced tests and student tasks											
	Standardized, norm-referenced tests						Student tasks					
	Language skills		TOPEL - phonol. awareness		TOPEL - print knowledge		Gift delay-wrap task		Gift delay-bow task		Bear/dragon task	
TEEM	0.040	0.05	0.490	0.10	1.077	0.10	−0.031	−0.05	−0.107	−0.03	0.041	0.01
	(0.248)		(0.450)		(0.976)		(0.055)		(0.231)		(0.308)	
PALS	−0.160	−0.18	−0.184	−0.04	1.531	0.14	0.160	0.19	0.200	0.06	0.262	0.09
	(0.428)		(0.430)		(0.676)		(0.074)		(0.230)		(0.310)	
TEEM × PALS interaction	−0.141	−0.16	−1.689	−0.33	0.384	0.03	0.001	0.01	0.635	0.18	−0.585	−0.21
	(0.856)		(0.858)		(1.356)		(0.028)		(0.458)		(0.619)	
	Panel C: Student outcomes based on parent and teacher surveys											
	Parent surveys				Teacher surveys							
	CBQ total score - Parent		SCBE-30 total score - Parent		CBQ total score - Teacher survey		School liking		School avoidance			
TEEM	−0.012	−0.02	0.026	0.05	−0.032	−0.04	0.141	0.23	0.142	0.24		
	(0.042)		(0.040)		(0.086)		(0.076)		(0.064)			
PALS	0.100	0.15	0.071	0.14	0.020	0.02	0.073	0.12	0.037	0.06		
	(0.044)		(0.040)		(0.062)		(0.046)		(0.042)			
TEEM × PALS interaction	0.003	0.01	0.096	0.18	−0.040	−0.05	−0.123	−0.20	−0.072	−0.12		
	(0.078)		(0.080)		(0.124)		(0.091)		(0.083)			

Notes: School avoidance is reverse coded (positive numbers indicate less school avoidance). See text for definition of abbreviations. Statistical tests have been adjusted for multiple comparisons using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995; Hochberg, 1988). p<.10; *p<.05.

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

do control group parents, an effect size of 0.59 SD. Overall, parents receiving PALS coaching gain 0.41 SD and 0.19 SD more than control group parents on outcomes assessed during shared book reading sessions and during free play activities, respectively (shown in row 1 of Table 2, panel B). These results are consistent with prior experimental evaluations of PALS, in which we find effect sizes on parent behaviors of between 0.17 and 0.61, across parent outcomes (Landry et al., 2008, 2011a).

We report the effects of the PALS intervention on student outcomes in the second row of each panel in Table 3. Seven of the 17 outcomes are statistically significant after correcting for multiple comparisons. Students in the PALS condition outperform their control group on all four measures taken during the book reading sessions. The first coefficient shown is for book reading engagement, which measures on a one to five scale the extent to which students show interest and involvement in book reading, initiate interactions with their parent, or show enthusiasm for the activity (see the Appendix for additional information about student outcomes). The coefficient for PALS implies that students in the PALS, no-TEEM condition increase their measure of book reading engagement from their baseline score by 0.32 more points than do students in the business-as-usual condition, an effect size of 0.27 SD. PALS students also increase their scores on the TOPEL print knowledge section by 1.53 more points than control group students, an effect size of 0.14 (mean prescores range from 92.7 to 96.3 and the control group SD is 15.5; see Table A7). The right side of panel B shows PALS students outperform control group students on one measure of self-regulation (the gift wrap task, 0.19 SD). As shown in panel C, PALS positively impacts students' attention focusing and impulsivity, as measured through parent surveys, but does not affect social competence, school liking, or avoidance, as measured through parent and teacher surveys.

Effects of Receiving both TEEM and PALS

The magnitude and direction of estimates of interaction effects support our hypothesis of synergistic effects; however, none of the interaction effects are statistically significant. These results are shown in the third row of each panel in Table 3. The first coefficient in Table 3 for example, book reading engagement, suggests that PALS students who are also in TEEM classrooms increase their scores on language use (on a one- to five-point scale) by 0.63 points more than the sum of the additional gains (over the control group) for students in TEEM or PALS only. Six of the 17 interaction terms are negative, implying that by some measures the programs may be redundant, although the negative estimates are also not statistically significant. In sum, we find some evidence that the family and teacher coaching interventions produced synergistic effects above what would be found from summing the effect of both programs, but point estimates are not statistically significant.

Assessing the Costs of TEEM and PALS

Results of the cost analysis for TEEM are shown in the left panel of Table 4, with greater detail provided in Table A9.[8] The total annual cost of TEEM is $96,698 at Site 1 and $41,056 at Site 2. While curricular materials represent over one-third of the costs of TEEM, the largest proportion of costs, 41 to 46 percent, results from the salaried work time of the coach. The coach at Site 1 works nearly full-time on

[8] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Table 4 Total and per-pupil yearly costs of implementing the TEEM and PALS interventions across two sites each.

	Cost of The Early Education Model (TEEM)		Cost of Play and Learning Strategies (PALS)	
	Site one	Site two	Site one	Site two
Personnel time	$58,050	$25,556	$14,612	$14,835
Professional Development for TEEM/PALS coach	$2,443	$637	$1,349	$678
Materials, equipment, and travel	$36,205	$14,864	$4,883	$4,196
Total yearly cost	$96,698	$41,056	$20,844	$19,709
Total classrooms	10	4	–	–
Cost per classroom	$9,670	$10,264	–	–
Total students per site	170	80	7	6
Cost per student	$569	$513	$2,978	$3,285
Average cost per student across two sites	$541		$3,131	

Notes: Greater detail for the cost analysis is included in Tables A9 and A10. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

TEEM and collaborates with approximately 10 teachers each year. The coach at Site 2 works part-time and collaborates with four teachers per year. The full-time coach at Site 1 increases the total cost, but because they collaborate with more teachers, the annual cost per student across the two sites is similar, $569 and $513, respectively. In our cost-effectiveness calculations, we use the average per-pupil cost across the two sites, which is $541.

The right panel of Table 4 shows that the total annual cost of PALS is substantially less than TEEM, $20,844 at Site 1 and $19,709 at Site 2 (described in greater detail in Table A10). However, the costs per-pupil of $2,978 at Site 1 and $3,285 at Site 2 are substantially more than TEEM. As with TEEM, the primary cost of implementing PALS is the salaried work time of the coach, representing 68 percent of the total yearly cost at Site 1 and 73 percent at Site 2. The key factor driving differences in the costs between TEEM and PALS is scale. In both interventions, a full-time coach can collaborate with between seven to 10 individuals (teachers or parents) per year. Because the TEEM intervention pairs coaches with teachers, and teachers oversee whole classrooms of 17 to 20 students, the cost of one full-time coach is spread over up to 200 students. In contrast, because PALS pairs coaches with individual parents, and therefore, individual children, one full-time PALS coach is spread over only between seven to 10 students. TEEM is a more resource-intensive intervention overall, as it requires teachers to be released during the day to work with the coach and involves a substantial amount of curricular materials. Yet, TEEM is less costly on a per-student basis because of the scale of students over which a full-time coach is spread. As a result, the average per-student cost of PALS across the two sites, $3,131, is 5.8 times greater than the cost per student of TEEM.

Comparing the Cost-Effectiveness of TEEM and PALS

Cost-Effectiveness for Student Outcomes

In Table 5, we report cost-effectiveness results for student outcomes. The cost-effectiveness ratio for PALS suggests that each SD increase in student outcomes

Table 5 Cost-effectiveness comparisons of the TEEM and PALS interventions.

	Effects	Cost	Cost-effectiveness ratio ($/SD)	Effectiveness-cost ratio (SD/$1,000)
TEEM	0.092	$541	–	–
PALS	0.124*	$3,131	$25,348	0.039
TEEM and PALS total effect	0.278	$3,672	–	–

Notes: *p<.05. The main effects of TEEM and the TEEM × PALS interaction effects are both positive, but statistically insignificant. The cost-effectiveness ratios (and effectiveness-cost ratios) for TEEM and the TEEM × PALS interaction are therefore not reported. Effects are the average of all 17 student outcomes (see Table 3). Effectiveness-cost ratios represent the estimated effect on student outcomes (in standard deviations) for each additional $1,000 per student per year.

would cost $32,067. In other words, a $1,000 increase in spending that is allocated to parent coaching would increase measured student outcomes by 0.039 SD. For context, extant literature suggests Head Start overall has an effectiveness-cost ratio of 0.023 SD/$1,000. We do not report the cost-effectiveness ratios for the TEEM intervention or the combined intervention of TEEM and PALS because these two treatment groups did not have statistically significant effects. However, point estimates suggest that the impacts of TEEM are only slightly smaller than PALS, while the cost of TEEM is less than one-fifth that of PALS. Instructional coaching alone (TEEM only) may be the most cost-effective of the three approaches, but the small sample size prevents us from conclusively determining the true impact of TEEM. Conversely, the combined intervention of TEEM and PALS has the largest impact on student outcomes, but because of the high cost of parent coaching, the combined intervention has the highest overall cost, and is unlikely to be the most cost-effective approach.

Sensitivity Analyses of Cost-Effectiveness Ratios

We conduct a Monte Carlo sensitivity analysis (Levin et al., 2017) to determine plausible ranges of the cost-effectiveness ratio for PALS. Following recommendations from Levin and Belfield (2015), we use confidence intervals of effects estimates and alternate assumptions related to cost to gauge the sensitivity of cost-effectiveness ratio estimates. We focus on the impact of PALS, since it is the only intervention with statistically significant effects on students. We assume the summary impact effect size estimate of PALS on students' cognitive and non-cognitive outcomes, 0.124 SD, is normally distributed and allow the effect to vary by the end points of the 97.5 percent confidence interval, which are 0.063 and 0.184. We allow coach salaries to range by 30 percent above and below the mean and allow the average number of parents with whom PALS coaches collaborate to range by plus or minus 30 percent of the total number of collaborating parents (plus or minus two parents).[9] The salary range reflects roughly double the range of the highest and lowest salaries reported in the Center for Benefit-Cost Studies of Education's Cost Tool Kit (2015) for individuals with similar qualifications as family coaches, while the number of parents that each PALS coach works with is based on the range of parent loads observed in our prior work (PALS coaches typically work with between six and 10 parents during a nine-month school year).

[9] We examined differences in the coach salary and the number of parents with whom PALS coaches collaborated because, as we discuss in later sections, these two factors were most important in determining the cost of PALS.

Allowing only the cost parameters to vary suggests a range of annual costs between $1,876 and $5,528 per student and a range of cost-effectiveness ratios of between $15,184 and $44,750 per SD increase in student outcomes (or between 0.022 and 0.066 SD/$1,000 per student). When we let both costs and effects vary, cost-effectiveness ratios range from $8,764 to $167,263 and effectiveness-cost ratios range from 0.006 to 0.114 SD/$1,000 per student.

DISCUSSION

This study examines whether providing coaching for both parents and teachers on the use of research-based effective parenting and teaching strategies would improve students' school readiness and whether these interventions would be cost-effective. In our discussion, we explore areas in which our initial hypotheses aligned with or differed from the results, examine how these interventions may change the overall cost-effectiveness of Head Start, consider strategies for improving the cost-effectiveness of TEEM and PALS, and discuss challenges and priorities for future cost-effectiveness research in early childhood education.

Cost-Effectiveness of Teacher and Parent Coaching

We hypothesized that the TEEM and PALS interventions would have positive impacts on measured teacher and parent outcomes and that these changes in teaching and parenting practices would positively impact student outcomes. We further hypothesized that students assigned to both conditions would see larger gains in outcomes, relative to the control group, than would be expected by summing the effects of TEEM and PALS. Finally, we noted that it was unclear *a priori* which of the three treatment conditions would be most cost-effective.

Analyses of teacher and parent outcomes generally align with our hypotheses: Results show that the interventions improve teacher and parent practices. Effects of PALS on parent behaviors are similar to previous randomized controlled trials (e.g., Landry et al., 2011a). Teacher effects for TEEM are also similar in direction to previous studies but have somewhat lower magnitudes. Both coaching interventions show positive effects on a range of student outcomes. Further, our results suggest that TEEM may improve the impact of PALS. Consistent with our initial hypotheses, students in the combined treatment condition have the largest effects, in some cases producing synergistic effects, although these results are not statistically significant. Similarly, while the effects of TEEM are not statistically significant, the low cost and moderate effects suggest that instructional coaching alone may be the most cost-effective of the three treatment conditions.

The general lack of statistically significant effects on student outcomes for the TEEM intervention conflicts with our prior research and thus warrants further discussion. We offer two plausible explanations for this inconsistency. First, while the effect size of 0.71 for teacher outcomes is large relative to findings from other experimental evaluations of teacher professional development (e.g., Garet et al., 2011; Garet et al., 2008), this effect is not as large as our prior evaluations of TEEM, in which we have found effect sizes ranging from 0.84 to 1.11 after one year of implementation and 1.71 after the second year (Landry et al., 2009, 2011b). A closer examination of the predicted scores (adjusted for teacher characteristics) suggests that treatment teachers do not experience large enough gains in their TBRS scores during the second half of the school year. As shown in Table A11, treatment teachers only score at the 3.0 level or above (on a four-point scale) on two of the 10 TBRS subscales. Relatedly, Table 2 shows that treatment teachers have statistically significant differences on most of the TBRS subscales at midyear assessments, but

only three of the 10 subscales by the end-of-year assessments. Figure A1 shows this trend visually.[10] While overall prescores are not statistically different from each other, treatment teachers have an average midyear score of 2.41 and control group teachers had an average score of 2.10, after adjusting for teacher characteristics. By the end of the year, treatment teachers increase their overall average score to 2.53 compared to 2.29 for control group teachers. In our prior work, we find positive and significant child outcomes when most TBRS subscales move to 2.80 or greater.

Second, the study may be underpowered, given our correction for multiple comparisons. Previous experiments involve approximately the same number of treatment conditions, but far larger samples (220 teachers in Landry et al., 2011b; and 262 teachers in Landry et al., 2009). Our power analyses conducted prior to implementing the experiment suggest that the design of the study, even after student attrition, is sufficiently powered to detect effect sizes for TEEM found in prior work. The power analysis shows that because TEEM is a classroom-level treatment, effect sizes of between 0.21 and 0.27 are necessary for statistical significance at conventional levels, depending on the *r*-squared for particular models and assuming standard normal distributions of outcomes and errors (see Schochet, 2008). However, the minimum detectable effect size is far larger after correcting for multiple tests (four of 17 outcomes are positive and significant prior to multiple comparison corrections; see Table A8). The PALS intervention, which shows a more robust pattern of positive student effects, has a larger sample size (and more degrees of freedom) because the intervention is student level, rather than classroom level. Given the lack of precision in our estimates of the effects of TEEM, we cannot rule out the possibility that with larger scaling, we might find significant results under multiple inference correction.

Last, we note that despite evidence that attrition is random, we cannot be certain that the 30 percent attrition rate for families did not bias our results. Under a 30 percent attrition rate, if those who leave the study all experience double the impact from the intervention as those who remain in the study, then the true impact of the intervention would be 30 percent higher and our estimates would be 23 percent below true impact.

Enhancing the Cost-Effectiveness of Head Start

Based on prior analyses of Head Start and the findings presented in this study, teacher and parent coaching models could not only increase the impact of Head Start, but also increase Head Start's overall cost-effectiveness. As part of our study, we monitor resource use in Head Start classrooms assigned to the control condition. Control group Head Start classrooms are not substantially different in terms of resource use from those included in the national sample for the Head Start Impact Study. We can therefore use estimates from Ludwig and Phillips (2008) and Puma et al. (2005) to estimate an incremental cost-effectiveness ratio (Levin & Belfield, 2015), which determines whether additional benefits of an add-on program improve the overall cost-effectiveness.

Ludwig and Phillips (2008) combine a summative measure of student outcomes estimated in the Head Start Impact Study (Puma et al., 2005) with an estimate of overall costs of Head Start and conclude that Head Start programs raise achievement by approximately 0.023 SD per $1,000 per child (an overall effect of 0.247 SD

[10] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

and cost of $10,517 in 2018 dollars). The effectiveness-cost ratios for PALS is 0.039 SD/$1,000 per pupil (an effect size of 0.124 SD and cost of $3,131; see Table 5). Therefore, the results of this study suggest that Head Start with PALS would have an overall effect of 0.371 SD (0.247 0.124), an overall cost of $13,648 ($10,517 $3,131) and an effectiveness-cost ratio of 0.027 SD/$1,000 per pupil, representing a 16 percent increase over the business-as-usual no-PALS Head Start condition. Although the effects are not statistically significant, the effectiveness-cost ratio for TEEM is 0.171 SD/$1,000 per pupil, implying that implementing instructional coaching in Head Start centers could increase cost-effectiveness by 30.7 percent. Finally, our results suggest that implementing both coaching models would increase total spending per student in Head Start centers to $14,189, but the total overall effect would increase to 0.525 SD, which would increase the overall cost-effectiveness of Head Start by 58 percent.

Factors Influencing the Cost of Early Childhood Interventions

Given the high cost of PALS relative to estimates of per-student spending in Head Start centers, we provide further discussion about how Head Start agencies and regional administrators could reduce the cost of parent or coaching interventions. Results of our cost analysis suggest that a primary determinant of the per-pupil cost of TEEM and PALS is the number of teachers and parents with whom coaches collaborate. This is consistent with prior work that emphasizes the importance of class size in determining educational expenditures (Knight, 2012; Krueger, 2003; Levin et al., 2017).

One approach to increasing the number of teachers and parents with whom coaches collaborate is to hold coaching sessions online through video conferencing. Virtual coaching decreases travel time, thereby reducing the direct costs of travel, and may allow coaches to collaborate with more teachers or parents (Kraft, Blazar, & Hogan, 2018). The direct costs of travel include mileage and salaried work time. Mileage costs alone accounted for 7 to 9 percent of total costs in TEEM and 13 to 15 percent in PALS, while the time cost of travel accounted for an additional 9 and 11 percent of total costs, respectively (see Tables A9 and A10 for more detailed cost data).[11] The geographic context in which these interventions are implemented plays a role in travel costs. Participating Head Start locations are spread across urban and suburban areas, requiring substantially more travel than if all Head Start centers are located within a smaller, more densely populated urban area. Indirect costs of travel include the loss of time that coaches could have spent collaborating with additional teachers or parents, thereby reducing per-student costs. Travel time could also be reallocated to allow for more in-depth coaching, which could potentially increase impact.

Our experimental design limited the number of collaborating teachers at each Head Start center. In order to prevent contamination across treatment groups, only one classroom per building was eligible to participate each year. However, in real-life settings, program implementers could benefit from economies of scale and positive spillovers if multiple teachers within the same center collaborate with coaches. Coaches would save travel time and potentially increase the number of collaborating teachers while teachers could learn not only from the coach, but from each other between coaching sessions. While these types of effects are not ideal in the case of an experiment, they may provide additional benefits in real-life scenarios that could result in cost savings or increased impacts.

[11] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

Future Application of Cost-Effectiveness Analysis in Early Childhood Settings

Most cost analyses of early childhood education compare entire preschool programs, such as Head Start, Abecedarian, or Perry Preschool, to either no preschool at all, or a business-as-usual condition in which families may or may not have access to early childhood education (Barnett, 1985; Belfield et al., 2006; Cattaneo, Titiunik, & Vazquez-Bare, 2017; Ludwig & Phillips, 2008). These types of studies are particularly amenable to cost-benefit analysis, which assesses whether the monetary social benefits of a policy warrant the social cost (Levin et al., 2017; Miller et al., 2018). That early childhood interventions provide social benefits that far outweigh the cost is widely established (Garces, Thomas, & Currie, 2002; Heckman et al., 2010). However, cost-benefit studies do not typically answer questions about the mechanisms for making existing programs more cost-effective. One of the priorities for future cost analysis research in early childhood education is the application of cost-effectiveness analysis to alternate interventions designed to enhance Head Start or other preschool programs.

Several insights from this study inform future cost-effectiveness research in early childhood education. First, as noted in Levin and Belfield (2015), collecting relevant cost data while the program is being implemented, rather than asking participants to make retrospective estimates of their time use allocations, provides more accurate data. Our approach in the current study includes time logs completed during the experiment and participants' retrospective estimates of time use. Participants report difficulty in estimating prior year time use patterns. Second, cost analyses provide insights into the causal mechanisms of program impacts. For example, parent coaches stress the importance of ongoing professional development, given the unique expertise required of their position. Finally, both cost-effectiveness researchers and early childhood program developers should consider the intensity of their interventions and expected magnitudes of outcomes prior to development and implementation. For instance, prior to implementation, program developers estimated that PALS would likely be orders of magnitude more costly than TEEM, which is consistent with our empirical findings. It may therefore be worth exploring whether a more intensive version of TEEM has a larger impact.

CONCLUSION

Providing universal prekindergarten is a common policy solution to addressing the nation's educational inequality. However, one of the lessons learned from the national Head Start Impact Study is that policymakers must pay close attention to the quality of programs. A key question is whether funding would be better spent by expanding Head Start programs to more families or by targeting additional resources in ways that improve the impact and cost-effectiveness of existing programs. The current study analyzes two interventions designed to enhance the efficacy of resources allocated to Head Start. Our results suggest coaching could improve the impact and the cost-effectiveness of Head Start programs. However, alternative interventions may be even more cost-effective and further research drawing on the tools of cost-effectiveness analysis is necessary. While early childhood interventions are one of the strongest approaches to improving educational opportunity for marginalized youth, promoting educational justice requires constant improvement of these programs and better use of limited resources.

DAVID S. KNIGHT is an Assistant Professor of Educational Finance and Policy and the Director of the Center for Education Research and Policy Studies at the University of Texas at El Paso, 500 W. University Avenue, El Paso, TX 79912 (e-mail: dsknight@utep.edu).

SUSAN LANDRY is the Albert and Margaret Alkek Distinguished Chair in Early Childhood and the Michael Matthew Knight Professor of Pediatrics at the University of Texas Health Sciences Center at Houston, 7000 Fannin Street, Suite 1800, Houston, TX 77030 (e-mail: susan.landry@uth.tmc.edu).

TRICIA A. ZUCKER is an Associate Professor of Pediatrics in the Children s Learning Institute at the University of Texas Health Sciences Center at Houston, 7000 Fannin Street, Suite 1800, Houston, TX 77030 (e-mail: tricia.zucker@uth.tmc.edu).

EMILY C. MERZ is an Associate Research Scientist at Teachers College, Columbia University, 525 West 120th Street, New York, NY 10027 (e-mail: em2154@tc.columbia.edu).

CATHY L. GUTTENTAG is an Associate Professor of Pediatrics in the Children s Learning Institute at the University of Texas Health Sciences Center at Houston, 7000 Fannin Street, Suite 1800, Houston, TX 77030 (e-mail: cathy.guttentag@uth.tmc.edu).

HEATHER B. TAYLOR is an Adjunct Associate Professor of Pediatrics in the Children s Learning Institute at the University of Texas Health Sciences Center at Houston, 7000 Fannin Street, Suite 2300, Houston, TX 77030 (e-mail: heather.taylor@uth.tmc.edu).

ACKNOWLEDGMENTS

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090212 to the University of Texas Health Science Center at Houston as well as a gift from the Still Water Foundation to the University of Texas Health Science Center at Houston. The research is also based on work supported by the National Science Foundation under Grant No. 1661097 and Grant No. 1740695. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education, the Still Water Foundation, or the National Science Foundation.

REFERENCES

Aos, S., & Pennucci, A. (2013). K–12 class size reductions and student outcomes: A review of the evidence and benefit-cost analysis. (Document No. 13–01–2201). Olympia, WA: Washington State Institute for Public Policy.

Bakermans-Kranenburg, M. J., Van Ijzendoorn, M. H., & Juffer, F. (2003). Less is more: Meta-analyses of sensitivity and attachment interventions in early childhood. Psychological Bulletin, 129, 195.

Barnett, W. S. (1985). Benefit-cost analysis of the Perry Preschool Program and its policy implications. Educational Evaluation and Policy Analysis, 7, 333–342.

Barnett, W. S. (2011). Effectiveness of early educational intervention. Science, 333, 975–978.

Barnett, W. S. (2007). Revving up Head Start: Lessons from recent research. Journal of Policy Analysis and Management, 26, 674–677.

Belfield, C. R., Nores, M., Barnett, S., & Schweinhart, L. (2006). The High/Scope Perry Preschool Program cost–benefit analysis using data from the age-40 follow-up. Journal of Human Resources, 41, 162–190.

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.

Bierman, K. L., Heinrichs, B. S., Welsh, J. A., Nix, R. L., & Gest, S. D. (2017). Enriching preschool classrooms and home visits with evidence-based programming: Sustained

benefits for low-income children. Journal of Child Psychology and Psychiatry, 58, 129–137.

Bierman, K. L., Welsh, J. A., Heinrichs, B. S., Nix, R. L., & Mathis, E. T. (2015). Helping Head Start parents promote their children's kindergarten adjustment: The Research-Based Developmentally Informed Parent program. Child Development, 86, 1877–1891.

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2018). Cost-benefit analysis: Concepts and practice (5th Ed). Upper Saddle River, NY: Prentice Hall.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), Handbook of child psychology: Theoretical models of human development (pp. 793–828). Hoboken, NJ, USA: John Wiley & Sons, Inc.

Brown, C. P. (2013). Reforming preschool to ready children for academic achievement: A case study of the impact of pre-k reform on the issue of school readiness. Early Education & Development, 24, 554–573.

Brownell, R. (2001). Expressive one-word picture vocabulary test—Spanish-English bilingual edition. Novato, CA: Academic Therapy Publications.

Burns, M. S., Donovan, M. S., & Bowman, B. T. (Eds.). (2000). Eager to learn: Educating our preschoolers. Washington, DC: National Academies Press.

Campbell, F. A., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. Science, 343, 1478–1485.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. Developmental Neuropsychology, 28, 595–616.

Cattaneo, M. D., Titiunik, R., & Vazquez-Bare, G. (2017). Comparing inference approaches for RD designs: A reexamination of the effect of Head Start on child mortality. Journal of Policy Analysis and Management, 36, 643–681.

Center for Benefit-Cost Studies of Education. (2015). CostOut—The CBCSE cost tool kit. New York, NY: Author. Retrieved from: https://www.cbcsecosttoolkit.org.

Chiang, H. S., Clark, M. A., & McConnell, S. (2017). Supplying disadvantaged schools with effective teachers: Experimental evidence on secondary math teachers from Teach for America. Journal of Policy Analysis and Management, 36, 97–125.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.

Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. The Journal of Primary Prevention, 28, 67–91.

Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. Journal of Policy Analysis and Management, 32, 692–717.

Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., Howes, C., & Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? Early Childhood Research Quarterly, 21, 174–195.

Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. American Economic Review, 92, 999–1012.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H. S., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). The impact of two professional development interventions on early reading instruction and achievement. NCEE 2008–4030. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2011). Middle school mathematics professional development impact study: Findings after the second year of implementation. NCEE

2011–4024. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. Psychological Bulletin, 134, 31–60.

Guttentag, C. L., Landry, S. H., Williams, J. M., Baggett, K. M., Noria, C. W., Borkowski, J. G., Swank, P. R., Farris, J. R., Crawford, A., Lanzi, R. G., Carta, J. J., Warren, S. F., & Ramey, S. L. (2014). My Baby and Me: Effects of an early comprehensive parenting intervention on at-risk mothers and their children. Developmental Psychology, 50, 1482–1497.

Harris, D. N. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. Educational Evaluation and Policy Analysis, 31, 3–29.

Heckman, J. J., & Kautz, T. (2013). Fostering and measuring skills: Interventions that improve character and cognition. No. w19656. Cambridge, MA: National Bureau of Economic Research.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. Journal of Public Economics, 94, 114–128.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75, 800–803.

King, J. A. (1994). Meeting the educational needs of at-risk students: A cost analysis of three models. Educational Evaluation and Policy Analysis, 16, 1–19.

Knight, D. S. (2012). Assessing the cost of instructional coaching. Journal of Education Finance, 38, 52–80.

Knight, D. S. (2017). Are high-poverty school districts disproportionately impacted by state funding cuts? School finance equity following the Great Recession. Journal of Education Finance, 43, 169–194.

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. Developmental Psychology, 36, 220–232.

Kraft, M. A. (2018). Interpreting effect sizes of education interventions. Brown University Working Paper. Retrieved January 2, 2019, from https://scholar.harvard.edu/files/mkraft/files/kraft_2018_interpreting_effect_sizes.pdf.

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.

Krueger, A. B. (2003). Economic considerations and class size. The Economic Journal, 113, F34–F63.

Ladd, G. W., & Dinella, L. M. (2009). Continuity and change in early school engagement: Predictive of children's achievement trajectories from first to eighth grade? Journal of Educational Psychology, 101, 190–206.

LaFreniere, P. J., & Dumas, J. E. (1996). Social competence and behavior evaluation in children ages 3 to 6 years: The short form (SCBE-30). Psychological Assessment, 8, 369–377.

Landry, S. H., Anthony, J. L., Swank, P. R., & Monsegue-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. Journal of Educational Psychology, 101, 448–465.

Landry, S. H., Smith, K. E., Swank, P. R., & Guttentag, C. L., (2008). A responsive parenting intervention: The optimal timing across early childhood for impacting maternal behaviors and child outcomes. Developmental Psychology, 44, 1335–1353.

Landry, S., Smith, K. E., Swank, P. R., Zucker, T., Crawford, A. D., & Solari, E. F. (2011a). The effects of a responsive parenting intervention on parent-child interactions during shared book reading? Developmental Psychology, 48, 366–392.

Landry, S., Swank, P., Anthony, J., & Assel, M. (2011b). An experimental study evaluating professional development activities within a state funded prekindergarten program. Reading and Writing, 24, 971–1010.

Landry, S. H., Zucker, T. A., Williams, J. M., Merz, E. C., Guttentag, C. L., & Taylor, H. B. (2017). Improving school readiness of high-risk preschoolers: Combining high quality instructional strategies with responsive training for teachers and parents. Early Childhood Research Quarterly, 40, 38–51.

Lee, J., Finn, J., & Lui, X. (2012). Time-indexed effect size for p-12 reading and math program evaluation. Paper presented at the 2012 meeting for the Society for Research on Educational Effectiveness, Washington, DC.

Lee, R., Zhai, F., Brooks-Gunn, J., Han, W. J., & Waldfogel, J. (2014). Head Start participation and school readiness: Evidence from the early childhood longitudinal study–birth cohort. Developmental Psychology, 50, 202.

Levin, H. M. (2001). Waiting for Godot: Cost-effectiveness analysis in education. In Richard Light (Ed.), Evaluations that surprise. San Francisco, CA: Jossey-Bass.

Levin, H. M., & Belfield, C. (2015). Guiding the development and use of cost-effectiveness analysis in education. Journal of Research on Educational Effectiveness, 8, 400–418.

Levin, H. M., Belfield, C., Hollands, F., Bowden, B. A., Cheng, H., Shand, R., Pan, Y., & Hanisch-Cerda, B. (2012). Cost-effectiveness analysis of interventions that improve high school completion. New York, NY: Center for Benefit-Cost Studies of Education, Teachers College, Columbia University.

Levin, H. M., Glass, G. V., & Meister, G. (1987). A cost-effectiveness analysis of computer-assisted instruction. Evaluation Review, 11, 50–72.

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2017). Economic evaluation in education: Cost-effectiveness and benefit-cost analysis (3rd Ed.). Thousand Oaks, CA: Sage Publications, Inc.

Li-Grining, C. P. (2007). Effortful control among low-income preschoolers in three cities: Stability, change, and individual differences. Developmental Psychology, 43, 208–221.

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. (2007). Test of preschool early literacy. Austin, TX: ProEd.

Lonigan, C. J., & Whitehurst, G. J. (1998). Relative efficacy of parent and teacher involvement in a shared-reading intervention for preschool children from low-income backgrounds. Early Childhood Research Quarterly, 13, 263–290.

Ludwig, J., & Phillips, D. A. (2008). Long-term effects of Head Start on low-income children. Annals of the New York Academy of Sciences, 1136, 257–268.

Mashburn, A. J., & Pianta, R. C. (2006). Social relationships and school readiness. Early Education and Development, 17, 151–176.

Mead, S. (2017). Happy birthday, Head Start. Washington, DC: U.S. News & World Report. Retrieved October 14, 2017, from https://www.usnews.com/opinion/knowledge-bank/articles/2017-05-18/after-52-years-head-start-still-helps-families-and-children.

Miller, T., Kosiewicz, H., Tanenbaum, C., Atchison, D., Knight, D. S., Ratway, B., Delhommer, S., & Levin, J. (2018). Dual-credit education programs in Texas: Phase II. Washington, D.C.: American Institutes for Research.

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. Developmental Science, 8, 74–87.

Obama, B. (2013). Remarks by the President in State of the Union Address. Washington, DC: The White House, Office of the Press Secretary.

Parrish, T. B. (1994). A cost analysis of alternative instructional models for limited English proficient students in California. Journal of Education Finance, 19, 256–278.

Puma, M., Bell, S. H., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). Third grade follow-up to the Head Start Impact Study. Final Report. OPRE Report

2012–45. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.

Puma, M., Bell, S., Cook, R., Heid, C., & Lopez, M. (2005). Head Start impact study: First year findings. OPRE Report 2005–22. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods, 2nd Ed. Thousand Oaks, CA: Sage Publications, Inc.

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' preacademic skills: Self-regulation as a mediating mechanism. Child Development, 82, 362–378.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001) Investigations of temperament at three to seven years: The children's behavior questionnaire. Child Development, 72, 1394–1408.

Rust, J., & Golombok, S. (2009). Modern psychometrics (3rd Ed). New York, NY: Routledge.

Schochet, P. Z. (2008). Statistical power for random assignment evaluations of education programs. Journal of Educational and Behavioral Statistics, 33, 62–87.

Sweet, M. A., & Appelbaum, M. I. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. Child Development, 75, 1435–1456.

U.S. Department of Health and Human Services. (2010). Head Start impact study. Final Report. Washington, DC: Author.

Van Zeijl, J., Mesman, J., Van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Juffer, F., Stolk, M. N., ... & Alink, L. R. (2006). Attachment-based intervention for enhancing sensitive discipline in mothers of 1-to 3-year-old children at risk for externalizing behavior problems: A randomized controlled trial. Journal of Consulting and Clinical Psychology, 74, 994.

Walters, C. R. (2015). Inputs in the production of early childhood human capital: Evidence from Head Start. American Economic Journal: Applied Economics, 7, 76–102.

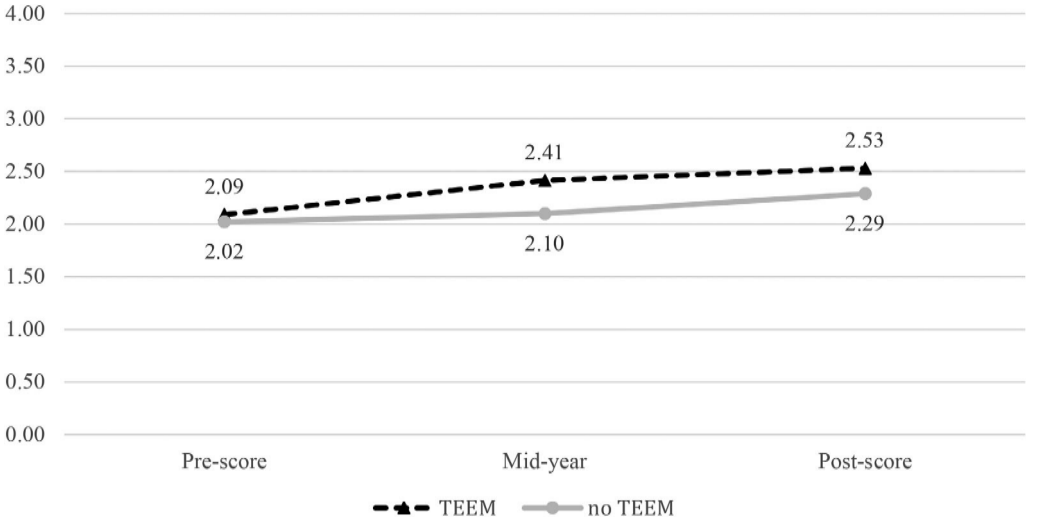
Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., & Fischel, J. E. (1994). A picture book reading intervention in day care and home for children from low-income families. Developmental Psychology, 30, 679.

Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. Journal of Policy Analysis and Management, 27, 122–154.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). Preschool language scale (4th Edition, Spanish Edition). San Antonio, TX: Harcourt Assessment.

APPENDIX A

ADDITIONAL FIGURES AND TABLES



Notes: The graph depicts the predicted values reported in Table A11. TBRS stands for Teacher Behavior Rating Scale and is measured on a scale from 0 to 4.

Figure A1 Overall TBRS Scores for Teachers Randomly Assigned to the TEEM and no-TEEM Treatment Conditions.

Table A1 Intervention session topics and schedule.

Timing	TEEM Session Topics	PALS Session Topics
October	1. TEEM Introduction[a]	PALS Introduction
November	2. Progress Monitoring Administration and Reports	Understanding Children's Signals
	3. Daily Schedule/Room Arrangement	Coaching - Signals
	4. Teaching Cycle/Gradual Release Model	Warm Responsiveness I
December	5. Classroom Management 1	Warm Responsiveness II
	6. Classroom Management 2	Coaching - Warm Responsiveness
January	7. Building Vocabulary 1	Guiding Children's Behavior I
	8. Building Vocabulary 2	Guiding Children's Behavior II
	9. Building Vocabulary 3	Coaching – Guiding Children's Behavior
February	10. Phonological Awareness 1	Review with Alternate Caregiver I
	11. Phonological Awareness 2	Reading with Young Children (includes coaching)
	12. Phonological Awareness 3	Maintaining Children's Interest
March	13. Using Data for Flexible Groupings	Coaching – Maintaining
	14. Letter Knowledge 1	Supporting Language Development – Labeling
	15. Letter Knowledge 2	Supporting Language Development – Linking Objects & Actions
April	16. Letter Knowledge 3	Coaching – Supporting Language Development
	17. Interactive Read Alouds 1	Review with Alternate Caregiver II
	18. Interactive Read Alouds 2	Using Responsive Behaviors during Daily Activities
May	19. Interactive Read Alouds 3/ Conclusion	Coaching – Everyday Activities, & Graduation
	20. Course Review and Teacher Self Reflection	

[a]This initial training was a two-day/14-hour training, whereas all other TEEM courses are of two-hour duration.

Table A2 Measure of fidelity of implementation of TEEM and PALS.

Domain of Implementation	Scale	*Mean (SD)*	Range
Measures of fidelity for TEEM			
TEEM coaching sessions	Total attended	12.87 (2.53)	9.00–19.00
TEEM coaching delivered	Total hours	33.61 (5.95)	22.00–45.25
TEEM coursework sessions	Total attended	14.51 (4.20)	0.00–19.00
Materials and environment setup	3-point rating scales from low to high fidelity	2.07 (0.49)	1.00–2.75
Cognitive instruction approaches	3-point rating scales from low to high fidelity	2.02 (0.50)	0.83–2.88
Responsive classroom mgmt.	3-point rating scales from low to high fidelity	2.15 (0.59)	1.00–3.00
Use of child prog. monitoring	Yes/No item, max of 1.00	40.0% (0.23)	0.00–0.94
Measures of fidelity for PALS			
PALS coaching sessions	Total completed	14.8 (6.38)	1.00–19.00
Length of PALS implementation	Weeks	23.6 (4.00)	15.00–34.00
Parent level of engagement	Average of 3 items, each on 3-point rating scale from low to high engagement in session	2.86 (0.32)	1.00–3.00
Parents' mastery of skills	Average of 2 items, each on 3-point rating scale from low to high mastery of skills with target child	2.54 (0.56)	1.00–3.00
Parents' generalization of skills	Average of 2 items, each on a 3-point scale from low to high mastery of skills with children other than target child	2.43 (0.61)	1.00–3.00
Parent ease of engagement	3-point rating scale from "very difficult to engage" to "easy to engage," reflecting amount of effort needed by coach to keep parent engaged in session	2.85 (0.38)	1.00–3.00

Table A3 Differences in characteristics of initial randomized student sample and final student sample after attrition.

	TEEM and PALS		TEEM, no PALS		PALS, no TEEM		Control		Total	
	Initial 156	Final 102	Initial 157	Final 112	Initial 155	Final 109	Initial 156	Final 111	Initial 623	Final 434
Average age in years	4.51	4.42	4.47	4.38	4.38	4.34	4.27	4.40	4.41	4.38
Female	51.9%	47.6%	54.0%	51.9%	53.0%	52.4%	46.2%	48.1%	51.2%	48.6%
African American	26.5%	24.8%	31.2%	31.8%	29.4%	30.3%	29.7%	29.1%	29.2%	29.1%
Latinx	69.8%	74.3%	63.9%	64.6%	72.3%	68.8%	74.5%	70.9%	70.1%	69.5%
Caucasian/White	0.9%	1.0%	3.8%	3.6%	0.0%	0.0%	0.0%	0.0%	1.2%	1.2%
Asian	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table A4 Teacher education and experience level by treatment and attrition status.

Panel A: Teacher education level

	HS/GED		CDA cred.		AA		BA		Some Grad.		Total
Original Sample	5	6.3%	19	24.1%	10	12.7%	34	43.0%	11	13.9%	79	100%
TEEM	4	10.0%	11	27.5%	5	12.5%	16	40.0%	4	10.0%	40	100%
No TEEM	1	2.6%	8	20.5%	5	12.8%	18	46.2%	7	17.9%	39	100%
TEEM stayers	4	11.4%	10	28.6%	4	11.4%	14	40.0%	3	8.6%	35	100%
No TEEM, stayers	0	0.0%	8	22.9%	5	14.3%	15	42.9%	7	20.0%	35	100%
TEEM leavers	0	0.0%	1	20.0%	1	20.0%	2	40.0%	1	20.0%	5	100%
No TEEM leavers	1	25.0%	0	0.0%	0	0.0%	3	75.0%	0	0.0%	4	100%
TEEM replace	0	0.0%	1	25.0%	1	25.0%	2	50.0%	0	0.0%	4	100%
No TEEM replace	0	0.0%	0	0.0%	1	33.3%	2	66.7%	0	0.0%	3	100%
Final sample	4	5.2%	19	24.7%	11	14.3%	33	42.9%	10	13.0%	77	100%
TEEM	4	10.3%	11	28.2%	5	12.8%	16	41.0%	3	7.7%	39	100%
No TEEM	0	0.0%	8	21.1%	6	15.8%	17	44.7%	7	18.4%	38	100%

Panel B: Teacher experience level

	TEEM	No TEEM	Total	Diff. between TEEM and no TEEM
Original Sample (n = 79)	11.44	10.59	11.02	0.85
Final Sample (n = 77)	11.29	10.65	10.97	0.64
Stayers (n = 70)	11.56	10.36	10.96	1.20
Leavers (n = 9)	9.22	12.63	10.74	−3.41
Replacements (n = 7)	8.91	14.04	11.11	−5.13

Notes: "replace" stands for replacement teachers. HS/GED, CDA cred., AA, BA, and Some Grad refer to a high school diploma or general equivalency degree, Child Development Associate credential, Associate's degree, Bachelor's degree, and some graduate school, respectively. Experience (shown in panel B) is based on teachers' self-reported number of years working as an educator either in a preschool or other educational setting. None of the differences between TEEM and no-TEEM shown in panel B, column 4, are statistically significant at $p<0.05$.

Table A5 Description of Teacher Behavior Rating Scale (TBRS) variables.

Scale	Description
Teacher sensitivity	Measures sensitive responsiveness, support for children's emerging autonomy and active engagement in activities, warmth and encouragement, and the absence of negativity
Classroom community	Measures classroom rules, routines, and organization as well as to what extent the classroom is arranged to support children's play and routines
Book reading behaviors	Assesses whether teacher exhibits behavior during book reading that supports children's emergent literacy skills (e.g., whether vocabulary words are discussed; children actively involved as teacher encourages them to make comments and ask questions; teacher asks children questions related to book content)
Oral language use with students	Assesses the extent to which the teacher speaks clearly and uses grammatically correct sentences, models expressing ideas in complete sentences, uses scaffolding language, poses questions and comments that support children's thinking, relates previously learned material or concepts to a classroom activity, and engages children in conversations that involve turn-taking
Phonological awareness	Seven types of phonological awareness activities (e.g., listening; sentence segmenting; rhyming; syllable blending and segmenting; alliteration; and phoneme blending, segmenting, and manipulating) were rated for quality and quantity
Print and letter knowledge	Measures the extent to which the teacher involves children in activities that support their acquisition of print knowledge (e.g., discusses print concepts such as that text contains letters, words, and sentences; a letter wall is used as an interactive teaching tool)
Math concepts	Assesses whether the teacher involves children in activities supporting early math (e.g., organized, hands-on math activities, such as number, arithmetic, space and geometry activities; uses math manipulatives; incorporates math into daily routines)
Written expression	Measures whether the teacher models writing, provides children with opportunities and materials to engage in writing
Team teaching	Assesses extent to which teacher and assistant work together so that children receive ongoing instruction in center activities, small group activities, and read-alouds; assistant is engaged with the children's learning and supports the lead teacher
Learning centers	Measures the quality and quantity of learning centers in the classroom (e.g., whether materials and activities follow the current theme and are linked to learning goals)
Lesson plans	Lesson plans are organized and connected with learning objectives, and are implemented, as reflect in classroom activities (e.g., give detailed explanations linking theme-related material to learning objectives)

Table A6 Description of observed parent and child variables during parent-child free play (panel A) and book reading (panel B) sessions.

Parent Scale	Description
Panel A: observed parent and child variables during parent-child free play sessions	
Panel A1: Parent outcomes that represent their own category based on factor analysis	
Prompts	Request from parent often in the form of a question or directive to do something with a toy
Praise	Parent reacts to child's behavior in a positive encouraging manner (e.g., good, wow, great job)
Verbal scaffolding	The extent to which the parent used rich language input with the child, such as labeling objects and actions and using higher level language (e.g., verbally linking concepts to one another)
Provision of labels	Parent provides the name of an object, activity, or event
Maintain attention	Parent talks about or makes a request related to child's current focus of attention
Redirect focus of attn.	Parent talks about or makes a request about an object or activity that is different from child's focus of attention
Negativity	Measured parental impatience, angry or harsh tone of voice, critical comments, and physical expressions of negativity
Panel A2: Parent outcomes that fall under the construct "responsive language and behavior support," based on factor analysis	
Warmth	Measured the degree to which the parent used a positive tone of voice, praised and encouraged the child, expressed physical affection toward the child, and exhibited acceptance of the child's needs and interests
Responsiveness and flexibility	Measured the extent to which the parent responded promptly and appropriately to the child's cues, followed the child's lead and pacing, and expanded on the child's play interests
Positive affect	Parental expression of positive affect through smiles, laughing, and facial animation
Panel A3: Student outcomes observed during parent-child free play sessions	
Social engagement	Child's social/communicative behaviors while interacting with the parent (e.g., gestures, verbal initiating, and responding to parent)
Cooperation	Measured children's ability to engage in play while also complying with parental requests
Panel B: observed parent and child variables during parent-child book reading sessions	
Panel B1: Parent outcomes that represent their own category, based on factor analysis	
Prompts	Request to say or do something from parent often in the form of directive or question
Praise	Statements by parent that offer encouragement or let the child know they are doing a good job
Verbal scaffolding	Questions, directives, or statements that give more information about words such as location, links with child experiences, etc.
Text duration time	Amount of time parent just reads the text
Tracking print	Parent's use of gestures to increase child's awareness of how books work (tracking text with finger, highlighting a word or photo)
Lang. building strat.	Labeling, verbal scaffolding, open prompts, using techniques to get child to verbalize
Negativity	Measured parental impatience, angry or harsh tone of voice, critical comments, and physical expressions of negativity

Table A6 Continued.

Parent Scale	Description
Panel B2: Parent outcomes that fall under the construct "responsive language and behavior support," based on factor analysis	
Enthusiasm and engagement	Measured parental use of positive talk and voice tone, praise and encouragement of the child, expression of interest in the child, attempts to make reading fun for the child, attempts to expand on the text presented in the book
Language comp. tech.	Measured parental engagement with the story in the book and attempts to help child engage and understand the text (e.g., adds dialogue related to text, acts out parts of the story, asks child questions about the story, helps child connect text with meaning)
Responsiveness and flexibility	Measured whether parent responded to child's cues/questions, matched pacing to the needs of the child, encouraged the child to be actively involved (e.g., turning pages), let child take the lead in storytelling
Panel B3: Student outcomes observed during parent-child book reading sessions	
Book reading engag.	Extent to which child shows interest and involvement in book reading, initiates interactions with parent, enthusiasm for activity
Language use	Use of language during book reading activity with parent, asks questions, communicates clearly with parent
Shared enjoyment	Laughter, smiling, warmth, cheerful tone, positive discussion, ease of interaction
Enthusiasm/initiative	Demonstration of excitement that results in initiating interactions with parent

Notes: Our factor analysis identifies three parent outcome measures taken during the parent-child free play sessions and three parent outcome measures taken during the parent-child book reading sessions that are highly correlated and collectively measured "responsive language and behavior support," as shown in panels A2 and B2. Panels A1 and B1 show all other parent outcomes, each of which measure independent constructs. Student outcomes assessed during parent-child free play sessions and parent-child book reading sessions are shown in panels A3 and B3, respectively.

Table A7 Descriptive statistics for student outcomes at pretest and posttest by intervention group.

	TEEM and PALS *M (SD)*		TEEM, no PALS *M (SD)*		PALS, no TEEM *M (SD)*		Control group *M (SD)*	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Language skills	−0.31 (0.91)	0.27 (0.87)	−0.08 (0.83)	0.45 (0.83)	−0.31 (.90)	0.24 (0.83)	−0.21 (0.09)	0.29 (0.85)
TOPEL–Print Knowledge	92.7 (13.2)	103.0 (13.8)	96.3 (15.8)	103.0 (14.0)	94.5 (14.8)	102.0 (15.1)	94.6 (15.5)	100.0 (14.4)
TOPEL–Phon. Awareness	81.9 (13.8)	84.1 (16.5)	83.8 (13.4)	88.2 (15.0)	81.2 (12.5)	84.2 (14.9)	83.3 (13.1)	84.3 (15.2)
CBQ Parent	4.19 (0.59)	4.36 (0.49)	4.16 (0.50)	4.29 (0.52)	4.23 (0.56)	4.42 (0.56)	4.14 (0.53)	4.27 (0.57)
CBQ Teacher	4.37 (0.80)	4.43 (0.90)	4.42 (0.85)	4.48 (0.96)	4.42 (0.86)	4.48 (0.76)	4.39 (0.80)	4.42 (0.81)
SCBE Parent	4.62 (0.49)	4.81 (0.52)	4.62 (0.50)	4.72 (0.51)	4.67 (0.53)	4.79 (0.48)	4.60 (0.55)	4.74 (0.52)
SCBE Teacher	4.70 (0.55)	4.84 (0.59)	4.68 (0.63)	4.85 (0.61)	4.70 (0.69)	4.76 (0.68)	4.70 (0.60)	4.81 (0.57)
School Liking	4.42 (0.52)	4.54 (0.49)	4.40 (0.67)	4.54 (0.52)	4.43 (0.63)	4.51 (0.52)	4.47 (0.55)	4.38 (0.60)
School Avoiding	1.39 (0.48)	1.33 (0.39)	1.36 (0.48)	1.32 (0.43)	1.49 (0.75)	1.40 (0.55)	1.39 (0.50)	1.47 (0.52)
Self-Restrain Gift	0.45 (0.66)	0.21 (0.40)	0.43 (0.60)	0.21 (0.37)	0.50 (0.62)	0.26 (0.45)	0.57 (0.71)	0.31 (0.46)
Self-Restrain Bow	0.29 (0.54)	0.29 (0.67)	0.39 (0.78)	0.22 (0.45)	0.37 (0.74)	0.24 (0.42)	0.45 (0.89)	0.32 (0.73)
Bear-Dragon	0.43 (0.50)	0.63 (0.49)	0.46 (0.50)	0.69 (0.47)	0.40 (0.49)	0.71 (0.46)	0.45 (0.50)	0.66 (0.47)
BR – Engagement	3.00 (1.13)	3.79 (1.09)	2.94 (1.24)	3.14 (1.17)	3.11 (1.17)	3.30 (1.29)	3.13 (1.11)	3.27 (1.17)
BR – Language Use	2.72 (1.24)	3.41 (1.17)	2.58 (1.34)	2.89 (1.32)	2.74 (1.26)	3.02 (1.49)	2.65 (1.23)	2.77 (1.37)
BR – Shared Enjoyment	2.80 (1.15)	3.44 (1.17)	2.69 (1.32)	2.83 (1.25)	2.96 (1.27)	3.04 (1.08)	2.98 (1.19)	2.83 (1.18)
BR – Initiative	8.52 (3.09)	10.63 (3.13)	8.22 (3.49)	8.85 (3.43)	8.81 (3.09)	9.36 (3.38)	8.76 (2.93)	8.87 (3.30)
FP – Cooperation	4.04 (1.01)	4.00 (1.01)	4.20 (0.78)	3.83 (1.15)	4.08 (0.95)	4.01 (0.92)	4.22 (0.96)	3.99 (1.08)
FP – Social Engagement	4.00 (0.94)	4.03 (0.92)	4.07 (0.93)	3.92 (1.04)	4.00 (0.95)	4.02 (0.96)	4.08 (0.91)	3.95 (1.09)

Notes: The "language skills" composite is comprised of the Expressive One-Word Picture Vocabulary Test, Preschool Language Scale – fourth edition, Auditory Comprehension and Expressive Communication scales, and TOPEL Definitional Vocabulary. Standard scores are provided for TOPEL Print Knowledge and Phonological Awareness. TOPEL, Test of Preschool Early Literacy; CBQ, Child Behavior Questionnaire; BR, Book Reading; FP, Free Play.

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

Table A8 Effects of TEEM and PALS on student outcomes, with correction for multiple statistical tests omitted (uncorrected version of Table 3) (*beta* coefficients, standard errors, and Cohen's *d*).

	Panel A: Student outcomes based on observational ratings of child behaviors											
	Book reading sessions								Free play sessions			
	Book reading engagement		Language use		Shared enjoyment		Enthusiasm/ initiative		Cooperation		Social engagement	
TEEM	0.206	0.18	0.248	0.20	0.265	0.21*	0.717	0.23	0.114	0.07	0.060	0.04
	(0.129)		(0.157)		(0.128)		(0.361)		(0.205)		(0.205)	
PALS	0.323	0.27*	0.350	0.28*	0.346	0.28**	1.009	0.32**	0.190	0.12	0.130	0.08
	(0.129)		(0.145)		(0.122)		(0.356)		(0.207)		(0.206)	
TEEM × PALS interaction	0.630	0.54*	0.298	0.24	0.398	0.32	1.322	0.42	0.294	0.18	0.039	0.02
	(0.258)		(0.289)		(0.241)		(0.706)		(0.411)		(0.410)	
	Panel B: Student outcomes based on standardized, norm-referenced tests and student tasks											
	Standardized, norm-referenced tests						Student tasks					
	Language skills		TOPEL - phonol. awareness		TOPEL - print knowledge		Gift delay-wrap task		Gift delay-bow task		Bear/dragon task	
TEEM	0.040	0.05	0.490	0.10	1.077	0.10	−0.031	−0.05	−0.107	−0.03	0.041	0.01
	(0.248)		(0.450)		(0.976)		(0.055)		(0.231)		(0.308)	
PALS	−0.160	−0.18	−0.184	−0.04	1.531	0.14*	0.160	0.19*	0.200	0.06	0.262	0.09
	(0.428)		(0.430)		(0.676)		(0.074)		(0.230)		(0.310)	
TEEM × PALS interaction	−0.141	−0.16	−1.689	−0.33*	0.384	0.03	0.001	0.01	0.635	0.18	−0.585	−0.21
	(0.856)		(0.858)		(1.356)		(0.028)		(0.458)		(0.619)	
	Panel C: Student outcomes based on parent and teacher surveys											
	Parent surveys				Teacher surveys							
	CBQ total score - Parent		SCBE-30 total score - Parent		CBQ total score - Teacher survey		School liking		School avoidance			
TEEM	−0.012	−0.02	0.026	0.05	−0.032	−0.04	0.141	0.23	0.142	0.24*		
	(0.042)		(0.040)		(0.086)		(0.076)		(0.064)			
PALS	0.100	0.15*	0.071	0.14	0.020	0.02	0.073	0.12	0.037	0.06		
	(0.044)		(0.040)		(0.062)		(0.046)		(0.042)			
TEEM × PALS interaction	0.003	0.01	0.096	0.18	−0.040	−0.05	−0.123	−0.20	−0.072	−0.12		
	(0.078)		(0.080)		(0.124)		(0.091)		(0.083)			

Notes: School avoidance is reverse coded (positive implies less school avoidance). See text for definition of abbreviations. $^{\dagger}p<0.10$; $^{*}p<0.05$; $^{**}p<0.01$.

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

Table A9 Total and per-pupil yearly cost of implementing the TEEM intervention as planned and for two sites (detailed version of Table 4).

		Prototype Model			Site one			Site two		
	Yearly salary/ annualized cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost
Panel A: Personnel time										
a TEEM coach	$56,250									
Facilitate initial coursework training		16	0.0077	$433	16	0.0077	$433	16	0.0077	$433
Facilitate coursework sessions		48	0.0231	$1,298	38	0.0183	$1,028	40	0.0192	$1,082
Travel to coursework sessions		10	0.0048	$270	10	0.0046	$257	5	0.0024	$135
Biweekly coaching meetings		480	0.2308	$12,981	380	0.1827	$10,276	158	0.0757	$4,259
Preparing for biweekly meetings		48	0.0231	$1,298	38	0.0183	$1,028	35	0.0168	$947
Travel to biweekly coaching meetings		360	0.1731	$9,736	285	0.1370	$7,707	105	0.0505	$2,840
Monthly training with coach supervisor		80	0.0385	$2,163	80	0.0385	$2,163	20	0.0096	$541
Other time allocated to coaching		605	0.2907	$16,352	636	0.3055	$17,186	324	0.1555	$8,748
b Coach supervisor	$93,750									
Review monthly reports from coaches		36	0.0173	$1,623	36	0.0173	$1,623	36	0.0173	$1,623
c Head Start teachers	$22,181									
Initial training for coursework		150	0.1038	$2,303	150	0.1038	$2,303	60	0.0415	$921
Time in coursework		400	0.2768	$6,140	400	0.2768	$6,140	160	0.1107	$2,456
Travel to coursework sessions		200	0.1384	$3,070	100	0.0692	$1,535	20	0.0138	$307
Coursework homework		200	0.1384	$3,070	100	0.0692	$1,535	60	0.0415	$921
Coaching during student time		400	0.2768	$0	380	0.2630	$0	150	0.1035	$0
Additional coaching during non-student time		0	0.0000	$0	95	0.0657	$1,458	8	0.0054	$121
Training for progress monitoring		16	0.0111	$0	16	0.0111	$0	16	0.0111	$0
Testing for progress monitoring		50	0.0346	$0	50	0.0346	$0	20	0.0138	$0
d Head Start directors	$25,172									
Meetings with coach to discuss TEEM		40	0.0277	$697	40	0.0277	$697	0	0.0000	$0
Ongoing meetings with directors			0.0000	$0	0	0.0000	$0	6	0.0040	$102
e Teacher assistants	$19,377									
Time in coursework		200	0.1384	$2,682	200	0.1384	$2,682	80	0.0033	$65
Coaching sessions			0.0000	$0	0	0.0000	$0	70	0.0029	$57

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

Table A9 Continued.

		Prototype Model			Site one			Site two		
	Yearly salary/ annualized cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost
Panel B: Professional development for TEEM coach										
a Teacher coach (personnel time outside school year)	$56,250									
Certification as a course trainer		4	0.0019	$108	4	0.0019	$108	4	0.0019	$108
Participate in coursework training		5	0.0024	$135	5	0.0024	$135	0	0.0000	$0
TSR summer institute		16	0.0077	$433	16	0.0077	$433	16	0.0077	$433
b Coach director	$106,250									
Review of practice sessions		4	0.0019	$204	4	0.0019	$204	0	0.0000	$0
Monthly training sessions		7	0.0035	$372	7	0.0035	$372	0	0.0000	$0
TSR summer institute		0.3	0.0002	$16	0.3	0.0002	$16	0	0.0000	$0
Prep. for TSR summer institute		0.8	0.0004	$41	0.8	0.0004	$41	0	0.0000	$0
c TEEM program director	$187,500									
Monthly training sessions		8	0.0055	$1,038	8	0.0055	$1,038	0	0.0000	$0
d Other resources for coach PD										
Materials for TSR summer institute	$50	1		$50	1		$50	1		$50
Coursework manuals for coach	$46	1		$46	1		$46	1		$46

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

Table A9 Continued.

		Prototype Model			Site one			Site two		
	Yearly salary/ annualized cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost
Panel C: Materials/equipment										
TEEM manuals for coaches and training videos	$200	1		$200	1		$200	1		$200
Hatch Positive Beginnings kit	$175	10		$1,750	10		$1,750	4		$700
Curriculum provided to HS program (both conditions)	$3,300	10		$0	2		$0	4		$0
School readiness inst. materials (Lakeshore "Ready to Read Toolkit")	$2,000	10		$20,000	10		$20,000	4		$8,000
Informational newsletters for parents (both conditions)	$5	20		$0	20		$0	8		$0
NetBook provided to teachers	$197	10		$1,967	10		$1,967	4		$787
Software licenses (Amplify/T-scape)	$200	10		$2,000	10		$2,000	4		$800
Coursework handouts for teachers (approx. half bilingual)	$32	10		$318	10		$318	4		$127
Classroom space for coursework sessions (donated)	$75	20		$1,500	20		$1,500	20		$1,500
Physical cost of coach travel (300 hours of driving)	$0.55	14,400		$7,920	11,400		$6,270	4,200		$2,310
Physical cost of teacher travel (200 hours of driving)	$0.55	8,000		$4,400	4,000		$2,200	800		$440
Total yearly cost				$106,613			$96,698			$41,056
Total classrooms		10			10			4		
Cost per classroom				$10,661			$9,670			$10,264
Total students		170		0	170			80		
Cost per student				$627			$569			$513

Note: The figures reported in this table are summarized in Table 4.

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

Table A10 Total and per-pupil yearly costs of implementing the PALS intervention as planned and for two sites (detailed version of Table 4).

		Prototype Model			Site one			Site two		
	Yearly salary/ annualized cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost
Panel A: Personnel time										
a Family coach	$35,625									
Initial meeting		2	0.0010	$34	2	0.0010	$34	2	0.0010	$34
Weekly coaching meetings		300	0.1442	$5,138	280	0.1346	$4,796	240	0.1154	$4,111
Preparing for weekly meetings		120	0.0577	$2,055	105	0.0505	$1,798	90	0.0433	$1,541
Travel to parent coaching sessions		240	0.1154	$4,111	140	0.0673	$2,398	120	0.0577	$2,055
Debriefing following a coaching session		360	0.1731	$6,166	210	0.1010	$3,597	180	0.0865	$3,083
Wkly mtg w/parent coach trainer/supervisor		30	0.0144	$514	30	0.0144	$514	30	0.0144	$514
Other time allocated to parent coaching time		595	0.2859	$10,185	65	0.0312	$1,113	183	0.0880	$3,134
b Coach supervisor	$106,250									
Weekly meetings with coaches		2	0.0007	$77	2	0.0007	$77	2	0.0007	$77
Other tasks related to supervision		2	0.0010	$102	2	0.0010	$102	2	0.0010	$102
c Assistant coach supervisor	$40,625									
Attend weekly meeting with coaches		2	0.0007	$29	2	0.0007	$29	2	0.0007	$29
Travel to and attend a coaching session		20	0.0097	$396	8	0.0038	$154	8	0.0038	$154
d Parents	n/a									
Weekly meetings		480	n/a	n/a	280	n/a	n/a	240	n/a	n/a
Independent time		240	n/a	n/a	140	n/a	n/a	120	n/a	n/a

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

Table A10 Continued.

		Prototype Model			Site one			Site two		
	Yearly salary/ annualized cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost	Hours/ units	FTE	Yearly Cost
Panel B: Professional development for family coach										
a Family coach (personnel time outside school year)	$35,625									
One-week course (lasts 3 years)		12	0.0059	$209	12	0.0059	$209	3	0.0015	$52
Practice sessions (lasts 3 years)		2	0.0009	$31	2	0.0009	$31	1	0.0004	$16
b Coach trainer	$106,250									
One week course (lasts 3 years)		1	0.0006	$62	1	0.0006	$62	1	0.0003	$31
Dev. of one week course (lasts 5 years)		1	0.0003	$35	1	0.0003	$35	1	0.0003	$35
Review practice sessions (lasts 3 years)		18	0.0088	$935	18	0.0088	$935	9	0.0044	$467
Weekly meeting with family coaches		2	0.0007	$77	2	0.0007	$77	2	0.0007	$77
Materials/equipment										
Coaching manuals	$25	1		$25	1		$25	1		$25
Coaching videos	$50	1		$50	1		$50	1		$50
Handouts for families	$2	12		$24	7		$14	6		$12
Camera with tripod (lasts 5 years)	$125	12		$1,501	7		$876	6		$750
Portable DVD player with screen (lasts 5 years)	$90	12		$1,077	7		$628	6		$538
Bag of toys	$30	12		$360	7		$210	6		$180
Cost of travel time for coach	$0.55	9,600		$5,280	5,600		$3,080	4,800		$2,640
Total yearly cost				$38,473			$20,844			$19,709
Total families		12			7			6		
Cost per student				$3,206			$2,978			$3,285

Note: The figures reported in this table are summarized in Table 5.

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

Table A11 Predicted scores on the Teacher Behavior Rating Scale (TBRS) at pretest, mid-year, and posttest, *beta* coefficients and Cohen's *d* effect sizes (detailed version of Table 2).

	No TEEM - M (SD)						TEEM - M (SD)						Pretest	Mid-year		Posttest	
	Pretest		Mid-year		Posttest		Pretest		Mid-year		Posttest		*b*	*b*	*d*	*b*	*d*
Total	2.0	(0.3)	2.1	(0.3)	2.3	(0.3)	2.1	(0.3)	2.4	(0.3)	2.5	(0.3)	0.07	0.31**	0.93	0.24*	0.71
TBRS subcategories																	
Book Reading Behaviors	2.0	(0.4)	1.9	(0.4)	2.1	(0.5)	2.0	(0.5)	2.4	(0.6)	2.3	(0.5)	0.12	0.42**	0.90	0.19	0.41
Learning Centers	2.1	(0.6)	2.2	(0.6)	2.6	(0.8)	2.2	(0.6)	2.5	(0.7)	3.0	(0.6)	0.11	0.36*	0.57	0.43*	0.68
Classroom Community	2.4	(0.6)	2.5	(0.5)	2.7	(0.5)	2.4	(0.6)	2.8	(0.5)	2.9	(0.5)	0.04	0.27*	0.45	0.23	0.38
Teacher Sensitivity	1.9	(0.3)	1.9	(0.4)	2.0	(0.4)	2.0	(0.4)	2.2	(0.4)	2.2	(0.5)	0.09	0.24*	0.68	0.16	0.45
Lesson Plans	2.0	(0.9)	2.6	(0.7)	2.6	(0.9)	2.2	(0.7)	2.9	(0.7)	3.0	(0.8)	0.23	0.38	0.47	0.38*	0.47
Math Concepts	1.9	(0.6)	1.8	(0.6)	1.8	(0.6)	1.8	(0.6)	2.2	(0.7)	2.1	(0.7)	−0.05	0.41*	0.67	0.20	0.32
Oral Language	2.1	(0.4)	2.1	(0.5)	2.3	(0.5)	2.2	(0.4)	2.4	(0.5)	2.5	(0.6)	0.1	0.29*	0.66	0.21	0.47
Phonological Awareness	2.6	(0.1)	2.6	(0.1)	2.6	(0.1)	2.6	(0.1)	2.6	(0.2)	2.7	(0.2)	–	–	–	–	–
Print and Letter Know.	2.0	(0.5)	2.2	(0.5)	2.3	(0.5)	1.9	(0.4)	2.4	(0.6)	2.6	(0.6)	−0.09	0.23	0.48	0.33*	0.70
Written Expression	1.7	(0.5)	1.8	(0.6)	1.9	(0.6)	1.8	(0.6)	2.0	(0.5)	2.1	(0.5)	0.12	0.20	0.38	0.20	0.38

Notes: Significance tests are corrected for multiple comparisons. The model predicting Phonological Awareness did not converge and results for this subscale are not reported in Table 2. $p<0.10$; *$p<0.05$; **$p<0.01$.

Journal of Policy Analysis and Management DOI: 10.1002/pam
Published on behalf of the Association for Public Policy Analysis and Management

APPENDIX B

ADDITIONAL INFORMATION ON STUDENT OUTCOME MEASURES

We assess a total of 17 separate student outcome measures at the beginning and end of each school year (results for each outcome are shown in Table 3). We track measures of (a) cognitive skills, (b) social and behavioral skills, (c) self-regulatory processes, and (d) executive function. These outcomes are based on school readiness research that we highlight in the main text of this article and align with the intended outcomes of the interventions. We use four different techniques to assess these student outcomes:

Observation ratings of child behavior conducted by researchers during child-parent book reading sessions and free play sessions (six outcomes labeled book reading engagement, shared enjoyment, enthusiasm/initiative, cooperation, social engagement, and language use);
Student "tasks" (three outcomes used to measure self-regulation and executive functioning);
Teacher and parent surveys (five outcomes used to assess social-emotional functioning, school liking, and school avoidance); and
Standardized, norm-referenced tests (three outcomes that assess language and literacy skills).

Observation Ratings of Child Behavior

Six outcome measures are assessed during observations of child behaviors, which take place during free play and book reading sessions (described in the main article). All observation scores are measured on a scale of one to five, with one indicating almost never and five indicating almost always. Child activities are videotaped for later coding by researchers (see the main text for coding and rater training procedures). During videotaped free play sessions, children are rated on previously validated scales (Guttentag et al., 2014; Landry et al., 2008) that measure social engagement and cooperation. We find the social engagement scale has an intraclass correlation coefficient (ICC) of 0.65 and for cooperation, ICC = 0.71, indicating that most of the variation in our scales is between observations, rather than between raters within an individual student observation. In the book reading sessions, students are rated on measures of book reading engagement (ICC = 0.88), language use (ICC = 0.89), shared enjoyment (ICC = 0.86), and enthusiasm/initiative (ICC = 0.79). These scales are detailed in Table A4 and have demonstrated high reliability and validity in previous research (Landry et al., 2011a).

Student Outcomes Related to Task Completion

We use three student tasks to assess students' self-regulation and executive function before and after intervention. Self-regulation is measured through a gift delay-wrap task (Kochanska, Murray, & Harlan, 2000; Li-Grining, 2007), in which children are told that they will receive a present but that they cannot turn and peek while the present is being noisily wrapped by the examiner for 60 seconds. Strategy scores (1 = leaves seat to peek, 2 = turns body in seat to peek, 3 = peeks over shoulder, 4 = does not peek) are given for every 15 seconds of the task and averaged to create a total strategy score for the task. Scores for strategy and number of seconds until first peek are strongly correlated (r = .76 at pretest; r = .78 at posttest) and are therefore standardized and averaged to create a total score (ICC = .94). A second test of self-regulation is the gift delay-bow task (Kochanska, Murray, & Harlan,

2000). The wrapped gift is placed in front of the child, who is told not to touch the gift until the examiner returned with a bow. The delay lasts two minutes. We create a total strategy score for the task based on the average of strategy scores measured every 15 seconds of the task. We measure strategy scores on a scale from 1 to 5 (e.g., 1 = removes toy and 5 = does not touch box). Scores for strategy and number of seconds until first touch are strongly correlated ($r = .62$ at pretest; $r = .63$ at posttest), and we therefore standardized and averaged these values to create a gift delay-bow total score. Finally, we measure changes in students' executive function using the bear/dragon task (Carlson, 2005). Children are told to follow the bear puppet's commands (e.g., touch your ear), but not the dragon puppet's commands. We use a pass/fail score based on recommendations from previous studies (a score of five out of six correct dragon trials constitutes passing the task; Carlson, 2005). For each of these three tasks, the interrater reliability is high, with ICC greater than 0.94 for each.

Student Outcomes Based on Teacher and Parent Surveys

We measure socio-emotional functioning and students' school liking and avoidance using surveys of parents and teachers, resulting in a total of five non-cognitive outcomes. Both parents and teachers complete the Children's Behavior Questionnaire (CBQ; Rothbart et al., 2001), which measures Attention Focusing, Inhibitory Control, and Impulsivity (40 items total). These categories include items such as *when picking up toys, usually keeps at the task until it is done* (Attention Focusing, 14 items), *can wait before entering into new activities if asked*, (Inhibitory Control, 13 items), and *usually rushes into an activity without thinking* (Impulsivity, 13 items). All measures are based on a scale from one (extremely untrue of the child) to seven (extremely true of the child). The CBQ scales demonstrate internal consistency in prior research ranging from 0.64 to 0.92 (average $\alpha = 0.77$) and interrater reliability estimates ranging from 0.67 to 0.92 (average $\alpha = 0.75$; Rothbart et al., 2001). We find in the current study that for the parent-rated CBQ total score, internal consistency (Cronbach's α) is 0.65 at pretest and 0.67 at posttest. For the teacher-rated CBQ total score, $\alpha = 0.81$ at pretest and 0.82 at posttest. The CBQ surveys result in two non-cognitive outcomes, the CBQ total parent score, and the CBQ total teacher score.

Parents and teachers also complete the Social Competence and Behavior Evaluation (SCBE-30; LaFreniere & Dumas, 1996). The SCBE contains three subscales with 10 items each: Social Competence, Anger/Aggression, and Anxiety/Withdrawal. We find that each of these measures has adequate internal consistency for the parent-rated variables and high internal consistency for teacher-rated variables (Cronbach's α are between 0.74 and 0.81 for parent ratings on pretests and posttests are between 0.80 and 0.93 for teacher-rated pre- and posttest). Finally, teachers complete the School Liking and Avoidance Questionnaire (SLAQ; Ladd & Dinella, 2009), in which teachers use a one (almost never) to five (almost always) scale to rate children on six items measuring school avoidance (e.g., complains about school) and seven items measuring school liking (e.g., enjoys most classroom activities). Internal consistency is $\alpha = 0.82$ to 0.85 at pretest and $\alpha = 0.78$ to 0.83 at posttest. Because teacher and parent surveys are administered in person, we receive a 100 percent response rate for all teachers and parents participating in the study.

Student Outcomes Based on Standardized Norm-Referenced Tests

We use three different assessments to measure students' language and literacy skills, which we categorize as cognitive outcomes. The Expressive One-Word Picture

Vocabulary Test (EOWPVT; Brownell, 2001) presents children with a series of illustrations depicting an object, action, or concept and they are asked to name each illustration. The Preschool Language Scale – Fourth Edition (PLS-4; Zimmerman et al., 2002) is a global oral language measure that includes two subscales that assess receptively understood language and the ability to verbally communicate with others. Last, the Test of Preschool Early Literacy (TOPEL; Lonigan et al., 2007) consists of three subtests: Print Knowledge (36 items), Phonological Awareness (27 items), and Definitional Vocabulary (35 items). As with all other student outcomes, we administer assessments before and after intervention each year. Due to high correlations among the language measures from the students' cognitive tests, we form a composite "language skills" factor consisting of the EOWPVT, both PLS subscales, and the Definitional Vocabulary subscale of the TOPEL. We find high internal consistency of this measure at both pretest (Cronbach's $\alpha = 0.93$) and posttest ($\alpha = 0.92$).

APPENDIX C

ADDITIONAL INFORMATION ON METHODS AND RESULTS OF THE COST ANALYSIS

Methods for Assessing the Cost of TEEM and PALS

We measure the total and per-pupil yearly cost of implementation of TEEM and PALS using the ingredients method (Levin et al., 2017). We first conduct interviews with the developers of the TEEM and PALS interventions and review program documents to identify the resources that are likely to be used for each intervention. Next, we assign dollar values for each resource based on their current market price, the prevailing salary for that employee (for coach trainers and supervisors), or, for some personnel, the average salary for an individual with similar qualifications. These average salaries are based on those included in the Center for Benefit-Cost Studies of Education Cost Tool Kit (2015). For physical resources that last multiple years such as cameras and DVD players, we annualize their value over their lifetime (five years in the case of technology resources). For start-up professional development that coaches would draw upon for multiple years, we average costs over the three years the intervention is implemented, discounting to present value costs using a 3 percent discount rate.

Methods for Assessing the Cost of TEEM

The cost analysis for the TEEM intervention is based on data collected through multiple interviews with all 39 teachers and both instructional coaches in each of the two sites (Houston and Austin). Interviews are followed with e-mail correspondence to ensure our estimates of resource use are accurate. The focus of interviews is to ascertain the total personnel time individuals allocate to the intervention as well as the quantity and quality of physical resources used to implement TEEM.[12] The interviews take place at the end of the third year of the study and coaches are asked to provide retrospective estimates of their time use and the time allocations of Head Start teachers and directors. These estimates are cross-checked through multiple sources including coach trainers, supervisors, and the program director as well as teacher and coach time logs. Time logs are collected as measures of implementation fidelity and track the estimated amount of time teachers and coaches spend collaborating as well as the total time teachers spend on CIRCLE coursework.[13]

[12] Physical resources required for implementing TEEM include, for example, instructional materials related to TEEM and the cost of travel to weekly classes for teachers and between Head Start classrooms for coaches. Any curricular or instructional materials provided to *both* treatment and control conditions (e.g., preschool curricula or informational newsletters) are not included as costs.

[13] Coaches' retrospective reports generally overestimate the amount of time allocated to (a) teacher time for coursework (20 two-hour sessions); (b) teacher time working with the coach (20 sessions lasting 2.5 hours each); and (c) coach time attending one-on-one coaching sessions, compared to the time logs collected at the time of the study. In cases where there were discrepancies, we confirm the accuracy of time logs with coach supervisors and correspond with coaches through e-mails. A central challenge of cost analysis of educational interventions is to decipher which time efforts resulted in additional costs, and which activities would take place in the absence of intervention (and, therefore, are not counted as costs because they do not result in resource re-allocation). For example, coaching that involves modeling and observing takes place while students are receiving instruction. Coaches' time during modeling and observing is included as a cost, but we exclude teachers' time because teaching lessons is part of their daily routine. In contrast, teachers have discretionary time during students' naptime to allocate to various work-related tasks. Coaches often use this time to reflect with teachers on their lessons and plan for next week's coaching session. These interactions are included as costs, represented as coaches' and teachers' salaried work time.

Coaches also report spending time with teacher assistants and Head Start directors. Head Start personnel time is not included in time logs, but this time is included as a personnel cost. Following the cost methods described in Levin and McEwan (2001) and in empirical analyses (e.g., Aos & Pennucci, 2013; Levin et al., 2012; Parrish, 1994), we present our final cost estimates as annual per-pupil costs.

Methods for Assessing the Cost of PALS

We estimate costs for the PALS intervention using the same methods described above, through document review and interviews with all stakeholders involved in implementation. We conduct three interviews with two of the family coaches and two interviews each with the coach trainer and coach supervisor (for a total of 10 interviews). All interviews are followed up with e-mail correspondence to fill in any missing data and confirm the accuracy of our analyses. We collect information on family coaches' salaried work time, personnel time devoted to supervision of coaches, all start-up and on-going professional development for coaches, and the total amount of parent time devoted to the intervention (reported by family coaches). We also collect data on all materials, equipment, and travel required as part of the PALS intervention. Each of these resources is assigned a dollar value using its market price, with the exception of parent time, which, following King (1994), was excluded from program costs. The per-pupil annual cost is calculated as the total annualized cost of all resources used for implementation divided by the number of students who received the intervention. As with the TEEM intervention, our final cost estimates are presented in yearly per-pupil figures.

Results of the Costs Analysis of TEEM and PALS

Costs of TEEM

Results of the cost analysis for TEEM are shown in the left panel of Table 4, with greater detail provided in Table A7. The per-pupil yearly cost of the TEEM intervention is slightly less than was suggested by program developers. Based on document analysis and interviews with administrators who regularly oversee implementation of TEEM, we find that the estimated cost of the intervention, prior to implementation for this study, is $10,661 per classroom or about $627 per student per year. Our empirical cost estimates from Site 1 and Site 2 suggest that the annual per-pupil cost is between $569 and $513, respectively (shown in the bottom row of Table A7). The coach at Site 1 collaborated with approximately 10 teachers each year and the coach at Site 2 collaborated with four teachers per year. Average class sizes across Sites 1 and 2 are approximately 17 and 20, respectively.

The primary cost of the TEEM intervention in both sites is the salaried work time of the instructional coach, representing 41 percent of the total and per-pupil yearly cost in Site 1 and 46 percent in Site 2. The annual salary for instructional coaches or "mentor teachers" in the TEEM intervention is $45,000 for nine and a half months and $56,250 when including 25 percent for fringe benefits. The coach at Site 1 allocated approximately 90 percent of her time each year to coaching during the nine and a half months of the intervention and her remaining salaried work time is allocated to research-related activities (her salary is partially paid through research grant funding). She, therefore, allocated a total of 71.3 percent of her yearly full-time equivalent (FTE) schedule to the TEEM intervention, at a cost of $44,531. The coach at Site 2 implemented TEEM over nine months each year and worked part-time, allocating a total of 33.8 percent of her yearly FTE to implementing TEEM at a cost of $18,984 (the sum of the TEEM coach costs are shown in panel A for Site 2 in Table A7). The coach supervisors review monthly reports for each coach and follow

up with weekly conversations with coaches and Head Start directors, which totaled approximately 36 hours per year per coach, at a total cost of $1,623 or about 2 to 4 percent of the total costs in each site (shown in Table A7 in panel A under Coach Supervisor costs).

Teacher time represents between 12 to 13 percent of the total yearly costs across sites. The coach at Site 1 collaborates with a total of 31 teachers over three years. On average, these teachers allocate a total of 1,291 hours per year to the TEEM intervention that could otherwise be used for other purposes. This time includes initial training, ongoing coursework, travel to weekly class sessions, homework associated with coaching sessions and coursework, training and testing for student progress monitoring, and meeting with the instructional coach during non-student time. The cost of teacher time does not include coaching sessions that take place while students are present in the classroom because this time is not considered a reallocation of the teacher's salaried work time. Although model developers anticipated that each teacher would spend one hour per week on homework, coaches in both sites report that teachers devote about half that amount. Head Start teachers are paid $15.35 per hour over approximately nine and a half months (1,445 hours) for an annual nine and a half-month salary of $22,181. Thus, the total cost of teacher time at Site 1 is $14,583, which represented approximately 13 percent of the total cost of the TEEM intervention. The coach at Site 2 collaborates with eight teachers over two years for a total of 494 hours (during non-student time), at a cost of $4,726, or 12 percent of total costs.

The majority of coaching sessions in both sites take place when students are in class, while coaches devote much less time to collaborating with teachers during student nap time or while a teaching assistant fills in. As a result, the cost of teachers' time associated with actual coaching is minimal compared to the overall cost of teacher time, most of which is allocated to coursework. This is consistent with cost analyses of instructional coaching in middle schools, in which the cost of teacher time represents less than 2 percent of the overall cost (Knight, 2012). Coaches also collaborate with Head Start directors and teaching assistants. The costs of teacher time, Head Start director time, and teaching assistant time are reported as the final three cost categories in panel A of Table A7.

The costs of professional development for the instructional coaches are reported in panel B of Table A7. Coach professional development in Site 1 is very similar to how the model developers planned it, in part because the intervention was implemented as part of an experiment and fidelity was important. Professional development costs include the instructional coach's salaried work time outside the regular nine and a half-month school year (certification in CIRCLE training and a yearly summer institute), salaried work time of the coach director and TEEM program director (ongoing monthly meetings with coaches), and additional materials and manuals for coaches. Because most of the instructional coach professional development takes place in group settings with other coaches, the salaried work time of the coach director and TEEM program director is minimal as it is spread over as many as 50 coaches (e.g., during the summer institute). The instructional coach's professional development at Site 2 is also similar to model developer estimates, except that she does not participate in monthly sessions with the coach director and the TEEM program director.

The TEEM intervention includes a large amount of materials and equipment as well as travel costs for both teaches and coaches, which we report in panel C of Table A7. Coaches are provided manuals describing the coaching process and training videos to use with teachers. Curricular materials, which alone comprised about one-fifth of the total costs across both sites, include school readiness instructional materials (Lakeshore "Ready to Read" toolkit) at a cost of $2,000 per classroom, a course software license ($200 per classroom), and additional materials to facilitate

lesson planning and coursework (Hatch Positive Beginnings kit is $175 per classroom and coursework handouts cost about $32 per classroom).

Head Start locations are spread across both urban and suburban settings of Austin and Houston, Texas. Coaches report that a substantial amount of their time was devoted to travel. Coaches drove approximately 45 minutes each way to meet with each teacher once per week at different Head Start centers, amounting to a total of 285 and 105 hours per year, for each coach, respectively (about 19 percent and 15 percent of total coaching time). Coaches also travel to CIRCLE coursework sessions, but classes were centrally located (all teachers meet in one location), so this travel represents less total time (see Table A7 for detailed cost information). While the coaches' travel time is included in their salaried work time, we also include the physical cost of travel valued at $0.55 per mile for a total of $6,270 at Site 1 and $2,310 at Site 2, or about 6 percent of total costs in each site. The physical costs of teacher travel to CIRCLE courses and the classroom space for these courses (donated by local school districts) are also included in the total cost.

Costs of PALS

We report annual costs of implementing the PALS intervention in the right panel of Table 4, with greater detail provided in Table A8. The per-pupil yearly cost of the PALS intervention is roughly in line with the cost suggested by program developers. We find that the estimated cost of the intervention, prior to implementation for this study, is $38,473 per family coach, or about $3,206 per student per year. The per-pupil annual cost is $2,978 in Site 1 and $3,285 in Site 2, which is shown in the bottom row of Table A8.

As with the TEEM intervention, the primary cost of implementing PALS in both sites is the salaried work time of the coach, representing 68 percent of the total yearly cost in Site 1 and 73 percent in Site 2. Family coaches in our study earn a yearly salary and fringe benefits totaling $35,625. While the PALS model is designed to take place over nine and a half months, coaches in Sites 1 and 2 completed the intervention in an average of eight and seven and a half months, respectively, each year of PALS implementation. During implementation, coaches worked full-time, but had additional responsibilities related to the research study (i.e., collecting consent forms, delivering newsletters to Head Start locations in the control group). The coach in Site 1 worked only 0.60 FTE and the coach in Site 2 worked 0.65 FTE, while the rest of their time was devoted to research-related activities. Thus, the total portion of their yearly salaried work time allocated to the PALS intervention was 40.0 percent FTE (for a cost of $14,250) and 40.6 percent FTE (for a cost of $14,473) at Sites 1 and 2, respectively. As a result of the reduced amount of time allocated to coaching, the family coach at Site 1 collaborated with seven families each year and the coach at Site 2 collaborated with six families per year, whereas model developers estimated that a full-time coach could collaborate with 12 families during a nine and a half-month academic year.

Supervision and professional development of coaches required approximately the same amount of time as program developers had intended. A coach supervisor and assistant coach supervisor attended weekly meetings with 10 family coaches and the coach supervisor checked in with family coaches weekly over the period of implementation. Model developers planned for the assistant coach supervisor to accompany family coaches on house visits once per month, but these supervisory visits took place every three months (as a result, the cost of personnel time for the assistant coach supervisor is slightly lower than planned). While the family coach at Site 1 received the same up-front professional development as planned, late hiring of the coach in Site 2 prevented her from receiving the full one-week course at the

beginning of the intervention, and she only completed one family coaching practice session (as opposed to two). The cost of professional development at Sites 1 and 2 represented 6 percent and 3 percent of total cost, respectively.

Finally, the cost of materials, equipment, and travel is similar to the intended implementation of PALS. The total cost of all material and equipment totals is approximately 8 percent of the total cost. Physical travel costs (as opposed to the personnel time associated with travel) represents about $3,000 in each site or about 13 to 15 percent of total costs. Under the prototypical PALS model (as planned prior to implementation) a family coach collaborating with 12 families is estimated to have physical travel costs over $5,000, which represents approximately 14 percent of total costs of the prototypical model.