CEPRS Working Paper | WP #2016-3

Cost-Effectiveness of Early Childhood Interventions to Enhance Head Start: Evidence from a Randomized Experiment

David S. Knight
Center for Education Research and Policy Studies
University of Texas at El Paso

Susan H. Landry
Tricia A. Zucker
Jeffrey M. Williams
Emily C. Merz
Cathy L. Guttentag
Heather B. Taylor
University of Texas Health Science Center at Houston

Center for Education Research and Policy Studies
College of Education, Suite 105
University of Texas at El Paso | 500 W. University Ave. | El Paso, TX 79968
(915) 747–5949
http://www.utep.edu/education/cerps/

Acknowledgements

The work is supported through funding from the College of Education and from the University of Texas at El Paso Office of the Provost.

Suggested Citation:

Knight, D. S., Landry, S. H., Zucker, T. A., Williams, J. M., Merz, E. C., Guttentag, C. L. & Taylor, H. B. (2016). *Cost-effectiveness of early childhood interventions to enhance Head Start: Evidence from a randomized experiment.* CERPS Working Paper 2016-3. University of Texas at El Paso, El Paso, TX.

Abstract

We evaluated the cost-effectiveness of instructional coaching and parent coaching models in Head Start using a randomized trial. The study design allowed us to compare the individual effects of each coaching model as well as their combined effect on student outcomes. Teachers receiving instructional coaching improved their use of language and literacy instructional practices, while parents receiving family coaching showed increases in numerous responsive parenting behaviors associated with positive child outcomes. Instructional coaching was more cost-effective than parent coaching in promoting these evidence-based practices. However, only the parent coaching model showed significant impacts on student outcomes. Parent coaching alone with no instructional coaching was therefore the most cost-effective of the three treatment conditions for improving student outcomes.

.

Cost-Effectiveness of Early Childhood Interventions to Enhance Head Start: Evidence from a Randomized Experiment

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; Dynarski, Hyman & Schanzenbach, 2013; Heckman, Moon, Pinto, Savelyev & Yavitz, 2010). Many policymakers have called for the provision of universal pre-kindergarten (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 expanding funding for Head Start often faces political challenges (Lachman, 2014). While many studies examine the effects of add-on preschool interventions, very 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.

The following study compared the cost-effectiveness of interventions that combine instructional coaching for preschool teachers with family coaching for the parents of students enrolled in preschool. We analyzed two widely implemented interventions, the Texas Early Educator Model (TEEM) and the Play and Learning Strategies (PALS) model. The design of our study allowed 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 family coaches with parents over the course of a school year.

We randomly assigned the TEEM intervention to teachers across 77 Head Start centers over a three-year period (2009-10 to 2011-12). Within classrooms, students were then randomly assigned to have their parents receive PALS. A total of 434 students were thus randomly assigned to both the TEEM and PALS interventions, just one of these interventions, or the control group. We tracked a wide array of pre-and post-intervention outcomes for teachers, parents, and students. Finally, we conducted a rigorous cost analysis of each intervention using the methods laid out in Levin and McEwan (2001). Our study addressed the following three 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 these three treatment conditions is the most cost-effective?

We found that both TEEM and PALS led to positive and significant improvements across a range of intermediate outcomes. The teacher coaching model showed substantially larger effect sizes on instructional practices than the parent coaching model had on measures of evidence-based parenting practices. Cost-effectiveness ratios suggest that TEEM was more cost-effective than PALS at increasing the use of evidence-based teacher and parent practices. Analyses of student outcomes showed that students assigned to the PALS only condition improved on several cognitive and non-cognitive outcomes, relative to the control group. However, the addition of TEEM instructional coaching to the PALS intervention did not significantly alter outcomes. Indeed, students assigned to TEEM only did not make significantly larger gains in outcomes

compared to a business-as-usual control group. Although implementation of PALS required far greater costs than the TEEM intervention, we conclude that PALS was the most cost-effective of the three treatment conditions tested. Based on prior analyses of the overall cost-effectiveness of Head Start (e.g., Puma et al., 2012; Ludwig & Phillips, 2008), implementing PALS in Head Start centers would increase the cost-effectiveness of Head Start by approximately seven percent.

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 produce any 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 the policy context, provide greater detail on the experimental design of our study, our analytic approach, and findings, and we conclude with further discussion and policy implications.

Background on Interventions to Enhance Preschool

A large body of research examines the effectiveness of preschool, which we review below. We then discuss prior research on early childhood interventions that draws on cost analysis, explore how these studies contribute to policy decision making, and explain how our study addresses gaps in this research. We provide greater detail on the TEEM and PALS interventions and review past research on their impacts on teachers, parents, and students. Finally, we offer a rationale for implementing these two interventions simultaneously and

present our hypotheses of their individual and synergistic effects on students.

The Impacts of Early Childhood Education

There is growing consensus among researchers that providing young children with high-quality classroom experiences prior to entering the K-12 public school system is one of the best approaches to narrowing educational opportunity and achievement gaps (e.g., Heckman et al., 2010). Research finds substantial positive social benefits of preschool programs that far outweigh the social costs (e.g., Barnett, 1985). For example, a randomized experiment of the High/Scope Perry Preschool Project involving 123 children in the 1960s found significant benefits of the program by age five that persisted well into adulthood (Schweinhart et al., 2005; Schweinhart & Weikart, 1997). A similar program, the Carolina Abecedarian Project has demonstrated positive long-term effects associated with intensive early childhood interventions including frequent home visits and full day preschool (Campbell & Ramey, 1995; Campbell et al., 2014).

Unfortunately, many of the existing state and federal preschool programs do not always provide high enough quality learning experiences to adequately prepare underprivileged students for kindergarten (e.g., Burchinal, Kainz & Cai, 2011; Wong, Cook, Barnett & Jung, 2008). In the three subsections below we discuss factors associated with preschool effectiveness, provide greater detail on definitions of school readiness, and describe a theoretical basis for providing complementary interventions at home and at school.

Exploring variation in preschool effectiveness. Prior research has identified significant variation in Head Start program effectiveness (Puma et al., 2012). Researchers have leveraged available data to identify the characteristics of high-quality programs (Garces, Thomas & Currie, 2002; Deming, 2009; Peck & Bell, 2014; Puma et al., 2012). Drawing on the analysis from the

Head Start Impact Study, Walters (2015) found that across 353 Head Start centers, full day service (as opposed to half day) and frequent home visits are most predictive of success, while other inputs typically thought to enhance preschool including High/Scope curriculum, teacher experience, and class size were unrelated to outcomes. However, two-thirds of the variation in Head Start effectiveness could not be explained with the available data. Importantly for our study, Walters did not have information on teachers' access to professional development or coaching, or on the quality of and practices used during home visits.

Other studies confirm that 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). Sweet and Appelbaum (2004) reviewed 60 home visit programs and found that while most significantly improved student outcomes, not all programs were 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., Bierman, Welsh, Heinrichs, Nix & Mathis, 2015). Several studies randomly assigned families to receive either home visits involving training on responsive parenting techniques or paper materials with similar information (e.g., Bakermans-Kranenburg, et al., 2003; Van Zeijl et al., 2006). Parents receiving home visits increased their use of responsive parenting techniques and their children showed gains in independent problem-solving, language, social and emotional skills, and behavioral development. Similarly, Bierman et al. (2015) found in a randomized controlled trial that home visits that provide parents with evidence-based learning games and guided pretend play to use with their children improved child literacy skills and academic performance prior to and during the kindergarten year.

At the same time, early childhood teachers also play an important role in contributing to

students' cognitive and social-emotional development (Burns, Donovan & Bowman, 2000). In summarizing experimental evidence of preschool programs, Barnett (2011) noted that highly effective programs – those that exceed the impacts of Head Start – staffed classrooms with teachers with a bachelor's degree and paid salaries equivalent to public-school teachers. Studies also link teacher training on specific pedagogical strategies or curricular programs with greater development of academic skills and social, behavioral, and self-regulatory processes (e.g., Raver et al., 2011). Students of teachers who were randomly assigned to training on the PATHS curriculum, for example, showed higher socio-emotional skills and were rated by parents and teachers as more socially competent compared to peers (Domitrovich, Cortes, & Greenberg, 2007). Although such studies do not parse out the effects of a new curriculum and teacher training associated with that curriculum, these studies suggest that instructional behaviors in pre-kindergarten settings impact student outcomes. Data from the Tennessee STAR experiment showed that the quality of instruction in kindergarten (as well as early elementary grades) has significant effects on students' achievement in those grades (Nye, Konstantopoulos & Hedges, 2004). Each of these studies demonstrate the important role of instructional effectiveness in promoting school readiness.

In summary, studies have highlighted significant benefits of early childhood interventions. While research demonstrates significant variation in preschool effectiveness, and many programs are found to be ineffective, instructional quality and effective home visits appear to be key factors contributing to success.

Defining and supporting school readiness. Research cited above draws on experimental and econometric analysis to identify effective programs or components of effective preschool interventions. Researchers in the field of child development have developed a parallel and

complimentary body of research that aims to define the indicators of school readiness that predict students' success and identify specific strategies to support school readiness (e.g, Lee, Zhai, Brooks-Gunn, Han & Waldfogel, 2014; Nobel, Norman & Farah, 2005). The broader domains of school readiness include (a) cognitive skills, (b) social and behavioral skills, (c) self-regulatory processes, and (d) executive function (Brown, 2013; Garon, Bryson & Smith, 2008; Mashburn & Pianta, 2006). Cognitive skills include oral language and early literacy skills such as letter knowledge and phonological awareness (Noble et al., 2005), while social and behavior skills include cooperation, attention focusing, and comfort with new situations. Finally, 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 et al., 2008).

Adults can support these particular indicators of children's school readiness by providing a stimulating environment with engaging playful learning activities (e.g., Burns et al., 2000; Sroufe, Coffino, & Carlson, 2010). Caregivers and teachers can encourage development of self-regulatory processes and social and behavioral skills by providing sensitive support for children's emerging attentional and communication skills and by helping children identify and achieve goals related to a particular task (e.g., Bernier et al., 2012). Responsive parenting is related to the development of social and self-regulatory skills, including those measured with executive function tasks as well as tasks that require inhibiting a response when problems are emotionally arousing (i.e., effortful control; Bernier, Carlson, Deschenes, & Matte-Gagne, 2012). Researchers have developed preschool interventions to support school readiness that focus on training teachers or parents to support children's cognitive skills, social and behavior skills, and executive function. However, no prior studies that we know of have examined the cost-

effectiveness of interventions aimed at improving instructional quality and parent responsiveness in particular.

Theoretical basis for providing complementary interventions. Despite substantial research on the value of parental involvement for youth (e.g., Crosnoe & Cooper, 2010; Graue, Clements, Reynolds, & Niles, 2004; Jeynes, 2012), few studies have experimentally examined the effects of "parent-plus-teacher" interventions (Lonigan & Whitehurst, 1998; Mol, Bus, de Jong, & Smeets, 2008; 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.

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

While few studies of preschool draw on the tools of cost-effectiveness analysis, many have used cost-benefit analysis to demonstrate the positive social returns to early childhood interventions (Barnett, 2007; 1985; 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 (Boardman, Greenberg, Vining, & Weimer, 2011). 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, number of high school dropouts averted, 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 & McEwan, 2001).

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., Levin, 2001; Yeh, 2010). Often, cost-effectiveness studies use standard deviations of achievement, or effect sizes, to measure outcomes. Based on the typical effects of educational interventions, Cohen (1988) suggested that effect sizes of 0.2, 0.5, and 0.8 might be considered small, medium, and large. However, these benchmarks ignore costs, and Harris (2009) suggests that any intervention that increases test scores by 0.025 standard deviations per $1,000 per student per year could be considered large. 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 standard deviations (Harris, 2009; Krueger, 2003; Levin, Glass & Meister, 1987; Yeh, 2010). While most analyses of Head Start and other preschool programs utilize cost-benefit analysis, Ludwig and Phillips (2008) summarize first-year findings of the National Head Start Impact Study (Puma et al., 2005), suggesting that Head Start had an average effect size for cognitive and non-cognitive outcomes of 0.247 (based on Table 1 of Ludwig and Phillips, 2008 and Exhibit 1 of Puma et al. 2005). The authors estimate that the cost of Head Start is approximately $10,284 per child (in 2016 dollars, although no formal cost analysis was

conducted) and thus conclude that the program raises achievement by approximately 0.024 standard deviations 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 0.024 standard deviations 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.

The Texas Early Education Model and the Play and Learning Strategies Intervention

One way to improve the impact of Head Start is by improving instruction (Barnett, 2011), which is the primary goal of the TEEM intervention. TEEM is an instructional coaching model aimed at improving teacher effectiveness in Head Start and other preschool programs. The TEEM intervention includes a package of teacher professional development resources including: (a) a two-day initial training, (b) in-class coaching, (c) coursework, (d) progress-monitoring, and (e) provision of instructional resources. 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 and interests, balancing teacher- and child-directed activities, providing contingent responses, and other instructional strategies.

The two-day initial training introduces teachers to the components of the TEEM framework and the topics covered in coursework and coaching sessions (see Appendix Table A1). The coursework takes place over approximately 20 bi-weekly two-hour sessions, typically led by the teachers' 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. Teachers are provided with laptop computers to ensure access to the web-based courses, complete “homework” activities, and administer progress-monitoring measures. The coach provides follow-up one-on-one two-hour coaching sessions every two weeks to facilitate implementation of instructional practices through side-by-side coaching (i.e., co-teaching) and modeling lessons. The coaches also spend 15-30 minutes on reflective follow-up conversations and reviews progress monitoring data. Last, the intervention provides supplemental kits including materials to support classroom management (HATCH “Positive Beginnings” kit) and school readiness instructional materials (Lakeshore “Ready to Read Toolkit”). In two large experimental studies, TEEM has demonstrated greater improvements in teachers’ practices, as measured by observation instruments, compared to teachers who did not receive TEEM. Prior experimental work also shows the TEEM intervention improves and children’s language and literacy skills compared to children in control classrooms (Authors, year; Authors, year).

Another potentially effective strategy for enhancing Head Start is improving the quantity and quality of home visits, which prior research has shown contributes to Head Start effectiveness (e.g., Walters, 2015). The PALS intervention provides a family coach who works one-on-one over the course of the school year (approximately 16-20 bi-weekly coaching sessions) to help parents implement research-based responsive parenting techniques. The PALS curriculum centers on a detailed manual and a set of training videotapes that the coach uses to help the parent learn to practice responsiveness behaviors. Responsive parenting behaviors include attending to and understanding young children’s communicative signals, following the child’s lead in play, 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.

Although PALS focuses on one designated parental guardian throughout the program, the child's other parent or another caregiver is welcome to attend sessions. Two sessions focus specifically on review and sharing of PALS skills with a second caregiver whom the participating parent invites to join the session. These sessions allow the coach to informally assess the parent's ability to articulate and demonstrate what she or he has been learning and familiarize the second caregiver with the target strategies. Several previous random-assignment efficacy studies have found significant positive effects of PALS on maternal responsiveness behaviors as well as children's language skills, cooperation, and social engagement (Author, year; Author, year).

In this study, we implemented the PALS and TEEM interventions on a parallel timeline. Both interventions target cognitive (literacy and language skills) and non-cognitive child outcomes (school liking, engagement, and avoidance and social-emotional and executive functioning). As such, we hypothesized that because lessons students learn in their preschool classroom are reinforced at home, implementing these interventions in unison may either provide 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. The only studies we know of that have tested the impacts of joint implementation of school and home interventions have found mixed results (Lonigan & Whitehurst, 1998; Whitehurst et al., 1994), and no such studies have examined cost implications.

Research Design and Analytic Approach

In this section, we first discuss the design of the experiment and the various types of data collected. We then review our experimental methods used to measure the causal effects of the two interventions. Finally, we describe our analytic approach for assessing costs.

Research Design

The study involved 77 Head Start classrooms in the Houston and Austin metropolitan areas over three years, from school year 2009-10 to 2011-12. Directors of six major Head Start agencies were contacted and all agreed to participate. We randomly assigned 39 classrooms to receive the TEEM intervention and assigned the other 38 classrooms to a business-as-usual control condition. Randomization was clustered by city to ensure an equal number of treatment classrooms were located in Houston and Austin. Parents of children in all classrooms were invited to participate and sign informed consents. Thus the population of causal inference is parents who enroll their children in Head Start and who are willing to implement the PALS intervention and participate in a study. Of the children with parental consent, six to eight children per classroom were randomly selected for participation. Half of these students (314) were randomly assigned to have their parents receive the PALS home-based parent program and the other half (309) were assigned to the business as usual control condition. The 2 x 2 design created 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 was recruited each year.

Teacher and student demographics are shown in Table 1. None of the differences in demographic characteristics were statistically significant; thus the randomization process was successful in creating similar groups for each treatment condition. Students were primarily Latina/o (69.5%) and African American (29.1%) and a small portion were White or Asian.

Approximately 40% of students in each condition spoke mostly Spanish at home, while all other students spoke a mix of English and Spanish (16.6%) or mostly English (44.0%). Most students were between three to five years old, with an average of 4.38. Parents reported that 96% of children qualified for free lunch. The majority of teachers were African American and Latina/o and teachers had 11 years of teaching experience on average. Approximately one-fifth of teachers delivered some of their lessons in Spanish and 6.6% used mostly Spanish.

The study involved 21 family coaches and two teacher coaches. We took significant measure 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 measured fidelity of implementation by tracking the extent to which teachers and parents participated in the various components of TEEM and PALS. Results of our fidelity measures are reported in Appendix Table A2. Senior research staff also supervised 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 monitored student and teacher attrition very closely. We initially recruited and administered pre-tests for 623 families and students and 434 are included in the final analysis (30.3% attrition rate). We found no meaningful or statistically significant differences in attrition across treatment conditions or in pre-test scores for those who left and those who remained in the study. We also had nine of the 79 teachers leave the study (11.4% attrition rate); however, we replaced six in time for pre-tests and a seventh early in the school year (we exclude this teacher in analyses of teacher effects). Teachers who left the study reported leaving for reasons unrelated to the intervention including classroom dissolution, transfer to a different classroom, or leaving the school. Pre-tests scores of the teacher observational ratings showed no

meaningful difference in those who left the study, replacement teachers, and those who remained in the study over the course of the year. In all cases, we recruited replacement teachers from the same Head Start center in which the previous teacher left. We used this same approach regardless of whether a teacher was in the TEEM or no TEEM condition

Methods for Assessing the Effects of TEEM and PALS

Our outcome measures for each intervention include both intermediate outcomes for teachers and parents as well as student outcomes.

Teacher and parent outcomes. To track instructional change associated with TEEM, we conducted classroom observations at the beginning, middle, and end of each school year for treatment and control teachers, using the Teacher Behavior Rating Scale (TBRS; Author, year). Classroom observers received certification in coding of TBRS, which includes demonstrating mastery of codes and achieving 80% reliability with coded videotaped lessons and side-by-side live coding with master coders. We also confirmed reliability of classroom observations by having two raters score on 15% of all classroom observations. The TBRS includes a total of 61 items that measure the quantity and quality of 11 subscales (e.g., Classroom Community, Oral Language, Book Reading Behaviors, etc; see Appendix Table A3 for greater detail). Each subscale is measured on a scale from zero to four. To measure changes over time on the TBRS, we used piecewise regression, which allows for different slopes between each observation period (also referred to spline regression; Akerhielm, 1999), and controlled for relevant teacher characteristics.

Intermediate outcomes for PALS were assessed through observer ratings of videotaped parent-child interactions during scheduled free play and book reading sessions (with toys and books provided by family coaches). Free play sessions involved parents monitoring their

children play with toys and book reading sessions involves parents reading books to their children. Parents were observed before and after implementation of the PALS intervention in their home by trained observers who were blind to the intervention status. Observers spent three weeks practicing coding, with the requirement that their ratings were 80% in agreement with a master coder. We held monthly meetings with observers to review and code videotapes as a group and included a second rater for approximately 15% of all parent observations. Observation instruments for book reading and free play sessions were developed by the authors that have been employed extensively in past studies (Authors, year). Instruments were found to be reliable and valid measures in evaluation of several 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 were factor analyzed to assess overlap in measures and in some cases, a single factor was identified from multiple constructs (e.g., enthusiasm / engagement, book comprehension technique, and responsiveness / flexibility all measured a single underlying factor referred to as "responsive language and behavioral support." Additional information on parent observation protocols is included in Appendix Table A4. We model the impact of PALS on parent outcomes using simple OLS regression, predicting end of year scores based on pre-tests and treatment status. Intermediate outcomes for both TEEM and PALS are reported in both raw coefficients and effect sizes (Cohen's *d*) so that comparisons can be made across measures of different scales.

Student outcomes. A total of 17 separate student outcome measures were assessed at the beginning and end of each school year (i.e., at the beginning and end of implementation of the interventions each year). We tracked measures of (a) cognitive skills; and non-cognitive skills including (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 school readiness research that we highlighted above. We assessed the 17 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
- Written and oral exams (three outcomes that assess language and literacy skills).

Initially, outcomes based on written and oral exams included of six total outcomes: three subsets of the Test of Preschool Early Literacy (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007), the Expressive One-Word Picture Vocabulary Test (EOWPVT; Brownell, 2001), and two subscales of the Preschool Language Scale – Fourth Edition (PLS-4; Zimmerman et al., 2002). Due to high correlations among the language measures, we formed a composite "language skills" factor consisting of the EOWPVT, both PLS subscales, and the Definitional Vocabulary subscale of the TOPEL.

For all observation measures we assessed intraclass correlations and we assessed Cronbach's alpha for all survey measures. These assessments demonstrated that our observation and survey measures were reliable and internally consistent, respectively, according to conventions established in prior literature (Rust & Golombok, 2009). We provide a more complete description of all student outcome measures, including our assessments of internal consistency and reliability in Appendix Table A4 and in the Online Appendix text. Appendix Table A5 provides descriptive statistics for student pre- and post-intervention outcomes. In short, we measured a total of 17 cognitive and non-cognitive student outcomes using four different data collection procedures.

Modeling student outcomes and interpreting coefficients. We estimated 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. 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:

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

Thus 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 interventions. 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, if γ_{21} is negative, we assume 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.

Given the multiple outcomes tested, we adjust all significance tests using the Benjamini-Hochberg method (Hochberg, 1988; Benjamini & Hochberg, 1995). We set alpha equal to 0.100 and conduct the Benjamini-Hochberg method separately for teacher, parent, and student

outcomes.[1] Before presenting our results, we first review our methods for assessing costs.

Methods for Assessing the Cost of TEEM and PALS

We measured the total and per-pupil yearly cost of implementation using the ingredients method (Levin & McEwan, 2001). Before collecting empirical cost estimates, we conducted interviews with the developers of the TEEM and PALS interventions and reviewed program documents related to expected resource use. Through this process, we identified all expected resources required of the intervention. We then assigned 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 were based on those presented in the Center for Benefit-Cost Studies in Education Cost Tool Kit (2015). For physical resources that last multiple years such as cameras and DVD players, we annualized 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 was implemented, discounting to present value costs using a 3% discount rate. This process allowed us to calculate the costs of each intervention as planned, prior to implementation (Levin, Catlin, & Elson, 2007). We refer to these costs as the prototype model costs. Once the prototype model costs for each intervention were established, we conducted an empirical cost analysis by conducting a series of interviews with study participants and cross-checking these data with study documents of research participants' time use.

[1] The Benjamini-Hochberg 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 significances 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.

TEEM. The cost analysis for the TEEM intervention was based on all 39 teachers and both instructional coaches in each of the two sites (Houston and Austin). We conducted a total of three interviews with each instructional coach, three interviews with the coach supervisor, two interviews with the coach trainer, and one interview with the program director. All interviews were followed with substantial email correspondence to ensure our estimates of resource use were accurate. The interviews took place at the end of the third year of the study and coaches were asked to provide retrospective estimates of their time use and the time allocations of Head Start teachers and directors. These estimates were cross-checked through multiple sources including coach trainers, supervisors, and the program director as well as teacher and coach time logs. Time logs were collected as measures of implementation fidelity and tracked the estimated amount of time teachers and coaches spent collaborating and the total time teachers spent on CIRCLE coursework.[2] Coaches also reported spending time with teacher assistants and Head Start directors. Head Start personnel time was not included in time logs, but this time is included as a personnel cost. Following the cost methods laid out in Levin and McEwan (2001) and in empirical analyses (e.g., Aos & Pennucci, 2013; Levin et al., 2007; Levin et al., 2012; Parrish, 1994) we present our final cost estimates as annual per-pupil costs.

The primary focus of interviews was to ascertain the total personnel time allocated to the intervention. A central challenge of this work was 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

[2] Coaches' retrospective reports generally overestimated 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 confirmed the accuracy of time logs with coach supervisors and corresponded with coaches through emails.

example, coaching that involved modeling and observing took place while students were receiving instruction. Coaches' time during modeling and observing was included as a cost, but we excluded 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 used this time to reflect with teachers on their lessons and plan for next week's coaching session. These interactions were included as costs for both coaches' and teachers' salaried work time. A second purpose of interviews was to gauge whether the quantity and quality of instructional materials and physical resources that were used to implement TEEM were similar to what program developers intended. This included, for example, 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) were not included as costs.

PALS. Costs for the PALS intervention were assessed using the same methods described above – using document review and interviews with all stakeholders involved in implementation. We conducted 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 were followed up with email correspondence to fill in any missing data and confirm the accuracy of our analyses. We collected 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 collected data on all materials, equipment, and travel used in the PALS intervention. Each of these resources was assessed with a dollar value using their market price, with the exception of parent time, which, following King (1994), was excluded from program costs. The per-pupil

annual cost was calculated as the total annualized cost of all resources used for implementation divided by the number of students that received the intervention. As with the TEEM intervention, our final cost estimates are presented in yearly per-pupil figures.

Methods for Comparing Cost-Effectiveness Ratios

We compare cost-effectiveness ratios for both intermediate and student outcomes. The overall effect size associated with each treatment condition is the average effect size across all measures. For example, the total intermediate effect of TEEM is the sum of the mid-year and end-of-year effect size (the end-of-year effect sizes measure change from mid-year to the end of the year). The total effect for PALS on parent outcomes is the average effect size of all parent outcomes. For purposes of making cost-effectiveness comparisons, we assume that a one standard deviation change in measures of teacher behavior is comparable to a one standard deviation change in measures of parent behavior. That is, we compare the annual per-student cost of altering measured teacher outcomes by one standard deviation to the annual per-student cost of altering measured parent outcomes by one standard deviation. Similarly, overall effects on student outcomes are the average effect sizes for all cognitive and non-cognitive outcomes. Cost-effectiveness ratios based on student outcomes provide insight into whether the intermediate outcomes ultimately affected students and which of the interventions is more cost-effective at improving student outcomes. 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.

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. Finally, for ease of interpretation, we report both the cost-effectiveness ratio (yearly per-student cost per standard deviation of effect size) for each treatment condition and the effectiveness-cost ratio, which measure the total effect size for each $1,000.

Results

Findings are presented in Tables 2-6. We discuss intervention effects, costs, and cost-effectiveness in the sections below.

Assessing the Effects of TEEM and PALS

Effects of TEEM. We found that the TEEM intervention led to observable instructional changes for teachers, but had fewer consistent and statistically significant effects on student outcomes. Intermediate effects of TEEM on teachers' instructional behavior are shown in Panel A of Table 2. Results are reported in both regression coefficients (subscales and the overall score on the TBRS are measured on a scale from zero to four) and standard deviation effect sizes (Cohen's *d*). Overall, teachers in the treatment group made larger gains on their mid-year assessment compared to the control group. Treatment teachers also outperformed those in the control group from the mid-year observation to the end-of-year observation; however, the relative gains of treatment teachers were larger in the first half of the school year, compared to the second half.

The first row of Panel A (Table 2) shows treatment effects for the overall score on the TBRS. By mid-year, treatment teachers increased their overall rating by 0.31 points more than control group teachers (on a 0-4 observation rating scale), an effect size of 0.93. By the end of the year, TEEM teachers gained an additional 0.24 points over the control group, for a total increase of 0.55 points (and a total effect size of 1.64 standard deviations from baseline). The

second row shows the results for the subscale book reading behavior, which measures whether the teacher supports children's emergent literacy skills during book reading sessions (e.g., whether vocabulary words are discussed, whether the teacher asks children questions related to the book content, etc.). Treatment teachers had an average increase of 0.42 points over control group teachers by mid-year and an additional 0.20 by the end of the year, corresponding to effect sizes of 0.90 and 0.41 standard deviations, respectively. The second subscale, Learning Centers (row 3), assesses whether the teacher arranges materials and class activities that (a) follow the current curricular theme and (b) are linked to learning goals. The coefficient of 0.36 suggests that treatment teachers improved their scores on the Learning Centers measure by over one-third more points than control group teachers, from pre-score to mid-year. The end-of-year coefficient for Learning Centers suggests treatment teachers made even larger gains during the second half of the year, relative to control group teachers (increasing their scores on this subscale by 0.43 more points than control group teachers). Teachers receiving the TEEM intervention outperformed control group teachers on almost every subscale from pre-test to the mid-year assessment and on several subscales from the mid-year assessment to the final post-test.

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, only four were positive and independently statistically significant, and none were statistically significant after correcting for multiple hypothesis tests. For example, we found that students whose teachers received the TEEM intervention gained an additional 0.27 points on the measure of shared enjoyment, assessed during parent book reading sessions. For context, the average pre-scores for shared enjoyment across the three treatment groups and the control group ranged from 2.69 to 2.98 out of a 5.00-point scale (based on

double-blind researcher observations using a previously validated instrument). This coefficient translates to an effect size of 0.21 standard deviations. Three other student outcomes were positive and independently statistically significant for TEEM (enthusiasm / initiative, school liking, and school avoidance), but not significant after correcting for multiple comparisons.

In sum, although the TEEM intervention led to significant changes in instruction, students assigned to TEEM classrooms did not experience consistently greater gains in outcomes compared to the business-as-usual control group. In our discussion, we provide several plausible explanations for the presence of positive teacher impacts and the general lack of positive and significant student outcomes.

Effects of PALS. Intermediate effects of the family coaching intervention are shown in Panel B of Table 2. We found that PALS positively impacted all of our observed parenting measures. The total scores for each parent activity observed (book reading and free play sessions) are based on the average of all effect sizes. Overall, teachers assigned to the PALS condition gained 0.41 standard deviations more from baseline measures than did control group parents in measures of parenting quality taken during shared book reading sessions (shown in row 1). Similarly, the final column of row 1 shows the average effect sizes of outcomes measured during free play activities (0.19 standard deviations). The overall score for PALS of 0.30 standard deviations is the average across all outcomes (not shown in Table 2, but used in cost-effectiveness comparisons with TEEM).

The first raw coefficient, labeled Prompts, suggests that PALS parents increased the number of times they prompted 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 increased their use of prompts by roughly a quarter of a standard deviation over

the control group parents (given the effect size of 0.24). The greatest impacts were seen on text duration time – a measure of the amount of time parents spent during book reading sessions just reading the text (as opposed to providing verbal scaffolding or tracking the print with the child). The coefficient of 23.9 suggests that parents in the PALS condition decreased their time spent just reading text to their child during the 10-minute book reading sessions by about 24 seconds more than did control group parents – an effect size of 0.59 standard deviations.

We report the effects of the PALS intervention on student outcomes in the second row of each Panel in Table 3. All but three of the 17 student outcomes showed positive gains for students whose parents worked with a PALS coach, and five of these outcomes were statistically significant after correcting for multiple comparisons. Students in the PALS condition outperformed 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 Appendix Table A4). The coefficient for PALS implies that students in the PALS, no TEEM condition increased their measure of book reading engagement from their baseline score by 0.32 more points than did students in the business-as-usual condition, an effect size of 0.27 standard deviations. PALS students also increased their scores on the TOPEL print knowledge section by 1.53 more points than control group students, an effect size of 0.14 (mean pre-scores ranged from 92.7 to 96.3 and the control group standard deviation was 15.5, online Appendix Table A5). The right side of Panel B and all of Panel C show that PALS had no impact on executive functioning (measured through student tasks) or attention focusing, impulsivity, 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, after correcting for multiple comparisons, none of the interaction effects are 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 were also in TEEM classrooms increased their scores on language use (on a 1 to 5-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. A total of 4 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, and the effect of combining programs is thus interpreted as compounding, but not synergistic or redundant. In short, we found little evidence that the family and teacher coaching interventions produced synergistic effects above what would be found from summing the effect of both programs, but we also did not find evidence that the programs were redundant.

Assessing the Costs of TEEM and PALS

Costs of TEEM. Results of the cost analysis for TEEM are shown in Table 4, with greater detail provide in Appendix Table A6. The per-pupil yearly cost of the TEEM intervention was slightly less than was suggested by program developers. Based on document analysis and interviews with administrators who regularly oversee implementation of TEEM, we found that the estimated cost of the intervention, prior to implementation for this study, was $10,661 per classroom or about $627 per student per year. Our empirical cost estimates from Site 1 and Site 2 suggest the annual per-pupil cost was between $569 and $513, respectively (shown in the bottom row of Table 4). The coach at Site 1 collaborated with approximately 10 teachers each year and the coach at Site 2 collaborated with 4 teachers per year. Average class sizes across Sites 1 and 2

were approximately 17 and 20, respectively.

The primary cost of the TEEM intervention in both sites was the salaried work time of the instructional coach, representing 41% of the total and per-pupil yearly cost in Site 1 and 46% in Site 2. The annual salary for instructional coaches or “mentor teachers” in the TEEM intervention was $45,000 for 9.5 months and $56,250 when including 25% for fringe benefits. The coach at Site 1 allocated approximately 90% of her time each year to coaching during the 9.5 months of the intervention and her remaining salaried work time was allocated to research related activates (her salary was partially paid through research grant funding). She therefore allocated a total of 71.3% 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 9 months each year and worked part time, allocating a total of 33.8% of her yearly FTE to implementing TEEM at a cost of $18,984 (shown in the first row of Table 4). The coach supervisors reviewed monthly reports for each coach and followed 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 2-4% of total costs in each site (shown in the second row of Table 4).

Teacher time represented between 12-13% of total yearly costs across sites. The coach at Site 1 collaborated with a total of 31 teachers over three years. On average, these teachers allocated a total of 1,291 hours per year to the TEEM intervention that could otherwise have been used for other purposes. This time included 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 did not include coaching sessions that took place while students were 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 reported that teachers devoted about half that amount. Head Start teachers were paid $15.35 per hour over approximately 9.5 months (1,445 hours) for an annual 9.5-month salary of $22,181. Thus the total cost of teacher time at Site 1 was $14,583, which represented approximately 13% of the total cost of the TEEM intervention. The coach at Site 2 collaborated with 8 teachers over two years for a total of 494 hours (during non-student time), at a cost of $4,726, or 12% of total costs.

The majority of coaching sessions in both sites took place when students were in class, while coaches devoted much less time to collaborating with teachers during student nap time or while a teaching assistant filled in. As a result, the cost of teachers' time associated with actual coaching was minimal compared to the overall cost of teacher time, most of which was allocated to coursework. This is consistent with cost analyses of instructional coaching in middle schools, in which the cost of teacher time represented less than 2% of the overall cost (Knight, 2010). Coaches also collaborated with Head Start directors and teaching assistants; the costs of teacher time, Head Start director time, and teaching assistant time are reported in the third, fourth, and fifth rows of Table 4.

The costs of professional development for the instructional coaches are reported in Panel B of Table 4. Coach professional development in Site 1 was very similar to how model developers planned, in part because the intervention was implemented as part of an experiment and fidelity was important. Professional development costs included the instructional coach's salaried work time outside the regular 9.5-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 took place in group settings with other coaches, the salaried work time of the coach director and TEEM program director was minimal as it was spread over as many as 50 coaches (e.g., during the summer institute). The instructional coach's professional development at Site 2 was also similar to model developer estimates, except that she did not participate in monthly sessions with the coach director and the TEEM program director.

The TEEM intervention included 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 4. Coaches were 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, included school readiness instructional materials (Lakeshore "Ready to Read" toolkit) at a cost of $2,000 per classroom, 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 were spread across both urban and suburban settings of Austin and Houston, Texas. Coaches reported 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% and 15% of total coaching time). Coaches also traveled to CIRCLE coursework sessions, but classes were centrally located (all teachers met in one location), so this travel represented less total time (see Appendix Table A6 for detailed cost information). While the coaches' travel time is included in their salaried work time, we also included 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% of total costs in each site. The physical cost of teacher travel to CIRCLE courses and the classroom space for these courses (donated by local school districts) were also included in the total cost.

Costs of PALS. We report annual costs of implementing the PALS intervention in Table 5, with greater detail provided in Appendix Table A7. The per-pupil yearly cost of the PALS intervention was roughly in line with the cost suggested by program developers. We found that the estimated cost of the intervention, prior to implementation for this study, was $38,473 per family coach, or about $3,206 per student per year. The per-pupil annual cost was $2,978 in Site 1 and $3,285 in Site 2, which is shown in the bottom row of Table 5.

As with the TEEM intervention, primary cost of implementing PALS in both sites was the salaried work time of the coach, representing 68% of the total yearly cost in Site 1 and 73% in Site 2. Family coaches in our study earned a yearly salary and fringe benefits totaling $35,625. While the PALS model is designed to take place over 9.5 months, coaches in Sites 1 and 2 completed the intervention in an average of 8 and 7.5 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% FTE (for a cost of $14,250) and 40.6% 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 7 families each year and the coach at Site 2 collaborated with 6 families per year, whereas model developers estimated that a full-time coach could

collaborate with 12 families during a 9.5-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 was 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 Site 1 and 2 represented 6% and 3% of total cost, respectively.

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

Comparing the Cost-Effectiveness of TEEM and PALS

Cost-effectiveness for intermediate outcomes. Panel A of Table 6 shows summative measures of effects, costs, and cost-effectiveness ratios for intermediate outcomes. As noted earlier, the total effect of TEEM on teaching behaviors is the sum of the total TRBS mid-year

and end-of-year score (1.64 standard deviations, SD). The total effect of PALS is the average effect sizes across all parent outcomes (0.30 SD). The effect size of TEEM is over five times as large as the effect size of PALS. Yet, the per-student yearly cost of TEEM is less than one-fifth that of PALS, suggesting that TEEM is far less costly to implement on a yearly per-student basis. The third column of Table 6 shows that each standard deviation increase in effective instructional practice costs $330 per student annually, while each standard deviation increase in responsive / effective parenting practices costs $10,328 per student annually. The final column of Table 6 provides effectiveness-cost ratios (the inverse of cost-effectiveness ratios, multiplied by 1,000). For each $1,000 increase in yearly per-pupil spending, TEEM improves measured teaching practices by 3.03 SD, whereas PALS improves measured parenting practices by 0.10 SD. These findings suggest that when implementing the TEEM and PALS interventions, improving parenting practices is far more costly than improving instructional quality by the same relative degree (e.g., by one standard deviation), compared to baseline conditions.

Cost-effectiveness for student outcomes. In Panel B of Table 6, we report cost-effectiveness results for student outcomes. Only the family coaching intervention had statistically significant effects on measured student outcomes. However, we show cost-effectiveness ratios for the non-statistically significant treatment conditions, TEEM and the TEEM x PALS interaction, for comparison purposes. The cost-effectiveness ratio for PALS suggests that each standard deviation increase in student outcomes would cost $32,067. In other words, a $1,000 increase in spending that is allocated to parent coaching would increase measured student outcomes would increase by 0.031 SD. For context, Head Start spends approximately $10,248 per student per year (Ludwig & Phillips, 2008) and Puma et al. (2005) estimate an effect size on cognitive and non-cognitive outcomes of 0.248 SD (implying an effectiveness-cost ratio of 0.024

SD / $1,000; Ludwig & Phillips, 2008). Although the cost and point estimates of effects suggested that both TEEM and the combination of TEEM and PALS would be more cost-effective at improving student outcomes, these effects are not statistically significant. The findings therefore suggest that PALS-only is the most cost-effective of treatments evaluated. Implementing TEEM in concert with PALS increases costs, but has no measurable effects on student outcomes.

Sensitivity analyses of cost-effectiveness ratios. We conduct a sensitivity analysis to determine plausible ranges of the cost-effectiveness ratio for PALS. Following recommendations from Levin and Belfield (2013), 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 on student outcomes because impacts on student outcomes are comparable to the broader literature and because PALS was the only intervention with statistically significant impacts. We assume the summary impact effect size estimate of PALS on students' cognitive and non-cognitive outcomes, 0.098 standard deviations, has a normal distribution with a 95% confidence interval ranging from 0.002 to 0.194 (based on the average t-statistic for PALS student outcomes). We allow coach salaries to range by 10% above and below the mean and we allow the average number of parents with which PALS coaches collaborated to range by plus or minus 2 parents.[3] The salary range reflects roughly the highest and lowest salaries reported in the Center for Benefit-Cost Studies in Education Cost Tool Kit (2015) for individuals with similar qualifications as the 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 6 and 10 parents during a 9-month school year).

[3] We examined differences in the coach salary and the number of parents with which 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 $2,218 and $4,875 per student and a range of cost-effectiveness ratios of between $22,712 and $41,446 per standard deviation increase in student outcomes (or between 0.024 and 0.044 standard deviations per $1,000 per student per year). When we let both costs and effects to vary around their estimated means, cost-effectiveness ratios range from $16,122 and $183,481 and effectiveness cost ratios range from 0.005 to 0.062 SD per $1,000 per student per year.[4]

Discussion

This study examined whether providing coaching for both parents and teachers on the use of research-based effective parenting and teaching strategies would improve students' school readiness. We focused particularly on the cost of these interventions, to examine whether their effects warranted the resources required for implementation. 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, explore 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 Parent and Teacher Coaching on Intermediate and Final Outcomes

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

[4] These estimates are based on a Monte Carlo simulation (Levin & McEwan, 2000) that assumes costs and effects are normally distributed around their mean. The mean, standard deviation and range for costs are $3,131, $247, and [$2,218, $4,875], respectively. The mean, standard deviation and 95% confidence interval for effects are 0.0976, 0.0219, and [0.0017, 0.1936], respectively. The "best- and worst-case scenarios" effectiveness-cost ratios range from 0.0004 to 0.0873 standard deviations per $1,000. These estimates are based on the maximum cost ($4,875) and the lower bound effect estimate, 0.0017, and the minimum cost ($2,218) and the upper bound effect estimate, 0.1936.

both conditions would see larger gains in outcomes, relative to the control group, than would be expected by summing the impacts of TEEM and PALS. Finally, we noted that it was unclear a priori which of the three treatment conditions would be most cost-effective, since PALS would likely have larger impacts, but larger costs, compared to TEEM. Analyses of intermediate outcomes aligned with our hypotheses: Both interventions improved teacher and parent practices. Relative to control groups, changes in our measures of effective teaching practices were generally larger than that of parenting practices. At the same time, TEEM had a lower annual cost per student and was therefore far more cost-effective at increasing evidence-based teaching than PALS was at increasing evidence parenting practices. However, only PALS showed significant impacts on student outcomes, and was therefore found to be more cost-effective in terms of positively impacting student outcomes. Consistent with several prior studies (e.g., Garet et al., 2011; Garet et al., 2008; Santagata, Kersting, Givvin & Stigler, 2011), the study demonstrates the difficulty in impacting student outcomes through teacher coaching, even when teachers appear to gain new knowledge and skills from coaching experiences.

The general lack of significant impacts from the TEEM intervention on student outcomes could result from the limited time TEEM was implemented (one academic year). In a prior evaluations of TEEM that included Head Start, child care, and public school pre-kindergarten teachers, we found significant positive effects for teachers after the first year of the intervention, while effects on child outcomes were not observed until the second year (Authors, year). Teachers were unable to implement new skills to the extent required to impact children's outcomes during the initial year. In the second year, teachers reported "making these skills their own," which likely translated to positive effects on children's learning.

An alternate explanation is that instruction did not improve fast enough during the school

year. Teachers in both the TEEM and non-TEEM conditions had TBRS scale pretest scores lower than those in our past studies (Authors, year). Although TEEM teachers made greater gains than control group teachers, TEEM teacher scores were not yet at the 3.0 level on a 4-point rating scale by mid-year and posttest. We have previously found positive and significant child outcomes when TBRS scores move to 2.8 or greater (Author, year). In short, steeper gains across the school year may be required to impact children's learning within the same school year, which may be difficult when teachers start at such low levels of instruction.

A third potential explanation for the lack of effects on student outcomes for TEEM – which is related to the moderate growth observed on TBRS scores – is that implementation of TEEM may have lacked sufficient fidelity. For example, across four observations of teachers' classrooms, observation scores averaged around two on a three-point scale of whether teachers set up their classrooms with TEEM materials, used cognitive instructional approaches, and had responsive classroom management. These measures of fidelity differ from the outcomes observed on the TBRS – the former assesses whether the teacher actually implemented the intervention, whereas the latter measures the consequences of that implementation (e.g., whether after implementing TEEM, the teacher exhibits behaviors that support children's emergent literacy skills). Although all teachers attended the initial startup professional development, 79.4% of the coursework sessions were completed across teachers on average. Additionally, we observed an average of only 40% child progress monitoring taking place (see Appendix Table A2). In short, higher fidelity to intervention may have increased TBRS scores, which in turn may have led to larger (and statistically significant) positive impacts for students assigned to TEEM classrooms.

We note that our power analyses conducted prior to implementing the experiment

suggested that the design of the study, even after the student attrition, was sufficiently powered to detect effect sizes for TEEM found in prior work (Authors, year). The power analysis showed 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 for more information on minimum detectible effects sizes). Although the minimum detectable effect size is larger after correcting for multiple tests, only 4 of 17 outcomes were significant prior to multiple comparison corrections.

Finally, cost estimates provide insight into the intensity of each coaching model, which may help further explain differences in the impacts of each intervention on student outcomes. TEEM is a relatively low-cost intervention, requiring roughly a five percent increase in per-pupil spending in Head Start.[5] Because TEEM matches between 8-10 teachers to an instructional coach, and each teacher has a class size of approximately 20 students, the student-to-coach ratio ranges from approximately 160 to 200 students per coach. PALS matches parent coaches to parents of individual students, implying a one-to-one ratio of coach to student. Although TEEM requires substantial curricular materials (representing about one-fifth of costs) and a significant amount of teacher time (12-13% of total costs), the larger student-to-coach ratio and lower overall per-student cost suggests that TEEM may not be considered as "intense" of an intervention, as measured by student-to-coach ratios and overall cost.

Enhancing the Cost-Effectiveness of Head Start

Based on prior analyses of Head Start described earlier and the findings presented in this study, the PALS intervention would not only increase the impact of Head Start, but also may

[5] Based on estimates of the cost of Head Start provided in Ludwig & Phillips (2008).

increase Head Start's overall cost-effectiveness. As part of our study, we monitored resource use in Head Start classrooms assigned to the control condition and concluded that these classrooms are not substantially different 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.

As described earlier, 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 0.024 SD per $1,000 per child (an overall effect of 0.248 SD and cost of $10,248 in 2016 dollars). The effectiveness-cost ratios for PALS is 0.031 SD / $1,000 per pupil (an effect size of 0.098 SD and cost of $3,131, see Table 6). Adding the PALS intervention would therefore increase the cost-effectiveness of Head Start to 0.026 SD / $1,000 per pupil, representing a 7% increase.

Factors Influencing the Cost of Early Childhood Interventions

One of the key findings of our study was that the higher cost intervention, PALS, was the only to one to impact student outcomes. We provide further discussion here about how Head Start agencies and regional administrators could reduce the per-pupil cost of both (or similar) coaching interventions.

A primary determinant of the per-pupil cost of both interventions is the number of teachers and parents with which coaches collaborate. This is consistent with past cost analyses, which show the central importance of class size in determining educational costs (Odden &

Picus, 2013; Krueger, 2003; Levin et al., 2007). One approach to increasing the number of teachers and parents with which coaches collaborate is to hold some or all of the one-on-one coaching sessions online through video conferencing, thereby reducing the amount of time allocated to travel. Our future work will explore the impact of electronic versions of TEEM and PALS in which coaches collaborate with teachers and parents through an online forum.

Coaches' travel to Head Start centers increased the total cost. The physical travel costs alone accounted for 7-9% of total costs in TEEM and 13-15% in PALS, while the time cost of travel accounted for an additional 9% and 11% of total costs, respectively (see Appendix Table A6 and A7 for more detailed cost accounting). The geographic context in which these interventions were implemented played a role in travel costs. Head Start locations were spread across urban and suburban areas, requiring substantially more travel than if all Head Start centers were located within a smaller, more densely populated urban area. This travel time could be also reallocated to more in-depth coaching (rather than expanding the number of teacher or parents that coaches work with), which could potentially increase impact (rather than lowering costs).

The number of collaborating teachers per Head Start center also influenced the cost. In order to prevent contamination across treatment groups, only one classroom per building was eligible to participate in the experiment 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 either cost savings or increased impacts.

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; Barnett & Masse, 2007; 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 & McEwan, 2000). That early childhood interventions provide social benefits that far outweigh the cost is widely established (Barnett, 2011; Heckman et al., 2010). However, cost-benefit studies do not typically answer questions about the mechanisms for making existing program more cost-effective. Cost-effectiveness analyses of alternate interventions aimed at improving Head Start or other preschool programs provide important information for policymakers and preschool administrators about strategies for increasing impact. Thus, one of the priorities for future cost analysis research in early childhood education is the application of cost-effectiveness analysis to interventions designed to enhance early childhood programs.

Several insights from this study inform future cost-effectiveness research in early childhood education. As recommended by Levin and Belfield (2013), we suggest collecting all relevant cost data while the program is being implemented, rather than asking participants to make retrospective estimates of their time use allocations. Our approach included both time logs completed during the experiment and participants' retrospective estimates of time use. Fortunately, we were able to confirm participant responses using the time logs, but participants reported difficulty in estimating prior year time use patterns. One benefit of conducting costs analyses along with impact studies is that during interviews, participants provide insights into the

mechanisms of program impacts. For example, coaches in our study reported that substantial time was lost due to travel. Similarly, parent coach trainers stressed the importance of training and ongoing professional development for parent coaches, given the unique expertise required of their position. Last, both cost-effectiveness researchers and program developers in early childhood settings should consider the intensity of their interventions and expected magnitudes of outcomes prior to development and implementation. For example, the study showed that, prior to implementation, program developers estimated that PALS would be about five times more costly than TEEM, and this estimate roughly matched our empirical findings. Given this information, it may be worth exploring the impact of a more intensive version of TEEM that reduces coaching loads, but could increase impact.

Conclusion

Providing universal pre-kindergarten is a common policy solution to addressing the nation's growing 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 that the parent coaching model 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 best approaches to improving educational opportunity for marginalized youth, promoting educational justice requires constant improvement of these programs.

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%	23.1%	29.1%
Hispanic/Latino	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%

Note: All teachers, parents, and students in the study identified as one of the four race/ethnicities listed. TEEM stands for Texas Early Education Model and PALS stands for Play and Learning Strategies. The control group did not receive the TEEM or the PALS intervention. Average parent characteristics were also similar across the PALS and no PALS conditions.

TABLE 2

Regression coefficients and standardized effect sizes (Cohen's d) for intermediate 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	0.93**	0.241	0.71*	Total	--	0.41**	--	0.19***
TBRS subcategories					*Parent observation score subcategories*				
Book read. behav.	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**

Note. For cost-effectiveness ratios, we use the sum of the mid-year and end-of-year gains on the Teacher Behavior Rating Scale (TBRS) to measure the overall effectiveness of TEEM on intermediate outcomes (1.64 standard deviations with a t-statistic of 3.47). We measure the overall effectiveness of PALS on intermediate outcomes by taking the average effect size across all measures, which is 0.303 with a t-statistic of 2.499. 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, 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).

TABLE 3

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

	Book reading engagement		Language use		Shared enjoyment		Enthusiasm / initiative		Cooperation		Social engagement	
	Panel A: Student outcomes based on observational ratings of child behaviors											
	Book reading sessions								Free play sessions			
TEEM	0.206 (0.129)	0.18	0.248 (0.157)	0.20	0.265 (0.128)	0.21	0.717 (0.361)	0.23	0.114 (0.205)	0.07	0.060 (0.205)	0.04
PALS	0.323 (0.129)	0.27+	0.350 (0.145)	0.28+	0.346 (0.122)	0.28+	1.009 (0.356)	0.32*	0.190 (0.207)	0.12	0.130 (0.206)	0.08
TEEM x PALS interaction	0.630 (0.258)	0.54	0.298 (0.289)	0.24	0.398 (0.241)	0.32	1.322 (0.706)	0.42	0.294 (0.411)	0.18	0.039 (0.410)	0.02
	Panel B: Student outcomes based on written exams and special tasks											
	Written exams						Students tasks					
	Language skills		TOPEL - phonol. awareness		TOPEL - print knowledge		Gift delay-wrap task		Gift delay-bow task		Bear/ dragon task	
TEEM	0.040 (0.248)	0.05	0.490 (0.450)	0.10	1.077 (0.976)	0.10	-0.117 (0.208)	-0.05	-0.107 (0.231)	-0.03	0.041 (0.308)	0.01
PALS	-0.160 (0.428)	-0.18	-0.184 (0.430)	-0.04	1.531 (0.676)	0.14+	-0.123 (0.207)	-0.06	0.200 (0.230)	0.06	0.262 (0.310)	0.09
TEEM x PALS interaction	-0.141 (0.856)	-0.16	-1.689 (0.858)	-0.33	0.384 (1.356)	0.03	0.015 (0.416)	0.01	0.635 (0.458)	0.18	0.585 (0.619)	0.21
	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 (0.042)	-0.02	0.026 (0.040)	0.05	-0.032 (0.086)	-0.04	0.141 (0.076)	0.23	0.142 (0.064)	0.24		
PALS	0.076 (0.039)	0.14	0.071 (0.040)	0.14	0.020 (0.062)	0.02	0.073 (0.046)	0.12	0.037 (0.042)	0.06		
TEEM x PALS interaction	0.003 (0.078)	0.01	0.096 (0.080)	0.18	-0.040 (0.124)	-0.05	-0.123 (0.091)	-0.20	-0.072 (0.083)	-0.12		

Note. 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.

TABLE 4

Total and per-pupil yearly costs of implementing the TEEM intervention as planned and for two sites

	Yearly salary / annualized cost	Prototype Model			Site one			Site two		
		Hours / units	FTE	Yearly Cost	Hours / units	FTE	Yearly Cost	Hours / units	FTE	Yearly Cost
Personnel time										
TEEM coach	$56,250	1,647	0.792	$44,531	1,482	0.713	$40,078	702	0.338	$18,984
Coach supervisor	$93,750	36	0.017	$1,623	36	0.017	$1,623	36	0.017	$1,623
Head Start teachers	$22,181	1,416	0.980	$14,583	1,291	0.893	$12,971	494	0.342	$4,726
Head Start directors	$25,172	40	0.028	$697	40	0.028	$697	6	0.004	$102
Teacher assistants	$19,377	200	0.138	$2,682	200	0.138	$2,682	150	0.006	$121
Professional development for TEEM coach										
Teacher coach	$56,250	25	0.012	$676	25	0.012	$676	20	0.010	$541
Coach director	$106,250	12	0.006	$633	12	0.006	$633	0	0.000	$0
TEEM program director	$187,500	8	0.006	$1,038	8	0.006	$1,038	0	0.000	$0
TSR summer inst. materials	$50	1	--	$50	1	--	$50	1	--	$50
CIRCLE manuals for coach	$46	1	--	$46	1	--	$46	1	--	$46
Materials, equipment, and travel										
TEEM manuals / training videos	$200	1	--	$200	1	--	$200	1	--	$200
Curricular materials	varies	10	--	$24,068	10	--	$24,068	4	--	$9,627
NetBook provided to teachers	$197	10	--	$1,967	10	--	$1,967	4	--	$787
Classroom space for eCIRCLE	$75	20	--	$1,500	20	--	$1,500	20	--	$1,500
Cost of coach travel (mi)	$0.55	14,400	--	$7,920	11,400	--	6,270	4,200	--	$2,310
Cost of teacher travel (mi)	$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			170			80		
Cost per student				$627			$569			$513

Note: FTE stands for full time equivalent. See text for detail on what is included in curricular materials.

TABLE 5

Total and per-pupil yearly costs of implementing the PALS intervention as planned and across two sites

	Yearly salary / annualized cost	**Prototype Model**			**Site one**			**Site two**		
		Hours / units	FTE	Yearly Cost	Hours / units	FTE	Yearly Cost	Hours / units	FTE	Yearly Cost
Personnel time (during the school year)										
Family coach	$35,625	1,647	0.792	$28,203	832	0.400	$14,250	845	0.406	$14,473
Coach supervisor	$106,250	4	0.002	$179	4	0.002	$179	4	0.002	$179
Asst. coach supervisor	$40,625	22	0.010	$425	9	0.005	$183	9	0.005	$183
Parents	n/a	720	n/a	n/a	420	n/a	n/a	360	n/a	n/a
Professional development for family coach										
Family coach	$35,625	14	0.007	$240	14	0.007	$240	4	0.002	$68
Coach trainer	$106,250	22	0.010	$1,109	22	0.010	$1,109	12	0.006	$611
Materials, equipment, and travel										
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	$125	12	--	$1,501	7	--	$876	6	--	$750
Portable DVD player	$90	12	--	$1,077	7	--	$628	6	--	$538
Bag of toys	$30	12	--	$360	7	--	$210	6	--	$180
Cost of coach travel	$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: FTE stands for full time equivalent.

TABLE 6

Cost-effectiveness comparisons of the TEEM and PALS interventions

	Effects	Cost	Cost-effectiveness ratio ($ / SD)	Effectiveness-cost ratio (SD / $1,000)
Panel A: intermediate outcomes				
TEEM	1.640*	$541	$330	3.031
PALS	0.303*	$3,131	$10,328	0.097
Panel B: student outcomes				
TEEM	0.078	$541	$5,934	0.169
PALS	0.098*	$3,131	$32,067	0.031
TEEM and PALS total effect	0.277	$3,672	$13,311	0.075

Note. * p<.05. For student outcomes, the main effects of TEEM and the TEEM x PALS interaction effects were both positive, but statistically insignificant. The cost-effectiveness ratios (and effectiveness-cost ratios) for TEEM and the TEEM x PALS interaction that are based on student outcomes (Panel B) are included for completeness, but should be interpreted as a statistically insignificant. The intermediate effect for TEEM is the total gain score on the Teacher Behavior Rating Scale and the intermediate effect for PALS is the average of all parent observation scores (see Table 2). Student effects for TEEM and PALS (Panel B) are the average of all 17 student outcomes (see Table 3). Effectiveness-cost ratios represent the estimated effect on intermediate or student outcomes (in standard deviations) for each additional $1,000 per student per year. Costs are measured as the average across each site, for each intervention (Tables 4 and 5).

References

Akerhielm, K. (1995). Does class size matter? Economics of Education Review, 14(3), 229-241.

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: metaanalyses of sensitivity and attachment interventions in early childhood. Psychological Bulletin, 129(2), 195.

Barnett, W. S. (2011). Effectiveness of early educational intervention. Science, 333(6045), 975-978.

Barnett, W. S. (2007). Revving up Head Start: Lessons from recent research. Journal of Policy Analysis and Management, 26(3), 674-677.

Barnett, W. S. (1985). Benefit-cost analysis of the Perry Preschool Program and its policy implications. Educational evaluation and policy analysis, 7(4), 333-342.

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(1), 289-300.

Bernier, A., Carlson, S. M., Deschênes, M., & Matte-Gagné, C. (2012). Social factors in the development of early executive functioning: A closer look at the caregiving environment. Developmental Science, 15(1), 12-24.

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(6), 1877-1891. doi: 10.1111/cdev.12448

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2006). Cost-benefit analysis: concepts and practice (4th Ed). Upper Saddle River, NY: Prentice Hall.

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(4), 554-573.

Brownell, R. (2001). Expressive One-Word Picture Vocabulary Test – Spanish-English Bilingual Edition. Novato, CA: Academic Therapy Publications.

Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), Quality measurement in early childhood settings (pp. 11-31). Baltimore, MD: Brookes.

Burns, M. S., Donovan, M. S., & Bowman, B. T. (Eds.). (2000). Eager to learn: Educating our preschoolers. Washington, D.C.: 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(6178), 1478-85. DOI: 10.1126/1248429.

Campbell, F. A., & Ramey, C. T. (1995). Cognitive and school outcomes for high-risk African-American students at middle adolescence: Positive effects of early intervention. American Educational Research Journal, 32, 743-772.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. Developmental Neuropsychology, 28(2), 595-616.

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

Lawrence Erlbaum.
Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. TESOL Quarterly, 14(2), 175-87.
Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. American Economic Journal: Applied Economics, 1(3), 111-134.
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(2), 67-91.
Dressler, C., & Kamil, M. L. (2006). First- and second-language literacy. In D. August & T. Shanahan (Eds.), Developing literacy in second-language learners. Mahwah, NJ: Lawrence Erlbaum.
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(4), 692-717.
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. Washington, D.C.: U.S. Department of Education, NCEE
Garet, M., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S. & Doolittle, F., (2010). Middle school mathematics professional development impact study: findings after the first year of implementation. Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, NCEE.
Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. American Economic Review, 92(4), 999-1012.
Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. Psychological Bulletin, 134(1), 31-60.
Harris, D. N. (2009). Toward policy-relevant benchmarks for interpreting effect sizes: Combining effects with costs. Educational Evaluation and Policy Analysis, 31(1), 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(1), 114-128.
Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75(4), 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), 1-19.
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(2), 220-232.
Krueger, A. B. (2003). Economic considerations and class size. The Economic Journal, 113(485), F34-F63.
Lachman, S. (2014, September 15). Here's how progressives plan to beat back the GOP tide. The Huffington Post. Retrieved October 10, 2015 from

http://www.huffingtonpost.com/2014/09/15/2014-election-progressives_n_5823904.html.
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(1), 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(4), 369-377.
Li-Grining, C. P. (2007). Effortful control among low-income preschoolers in three cities: Stability, change, and individual differences. Developmental Psychology, 43(1), 208-221.
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(1), 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(3), 400-418, DOI: 10.1080/19345747.2014.915604
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. Center for Benefit-Cost Studies of Education, Teachers College, Columbia University. Retrieved on January 4, 2014 from http://www.ny.frb.org/research/education_seminar_series/IESHighSchoolCompletion.pdf
Levin, H. M., Catlin, D., & Elson, A. (2007). Costs of implementing adolescent literacy programs. In D. Deshler, A. S. Palincsar, G. Biancarosa, & M. Nair (Eds.), Informed choices for struggling adolescent readers: A research-based guide to instructional programs and practices (pp. 61–91). Newark, DE: International Reading Association.
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. (2001). Cost-effectiveness analysis: Methods and applications (2nd Ed.). Thousand Oaks, CA: Sage Publications, Inc.
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(2), 263-290.
Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. (2007). Test of Preschool Early Literacy. Austin, TX: ProEd.
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(1), 257–268.
Ludwig, J., & Phillips, D. A. (2007). The benefits and costs of Head Start (Working paper No. w12973). Cambridge, MA: National Bureau of Economic Research.
Mashburn, A. J., & Pianta, R. C. (2006). Social relationships and school readiness. Early Education and Development, 17(1), 151-176.
Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. Developmental Science, 8(1), 74-87.
Nye, B., Konstantopoulos, S. & Hedges, L. V. (2004). How large are teacher effects? Educational Evaluation and Policy Analysis, 26(3), 237–57.
Obama, B. (2013). The White House, Office of the Press Secretary. Remarks by the President in state of the union address.

Odden, A. R. & Picus, L. O. (2013). School finance: A policy perspective (5th Ed.). New York, NY: McGraw-Hill.

Parrish, T. B. (1994). A cost analysis of alternative instructional models for limited English proficient students in California. Journal of Education Finance, 19(3), 256-278.

Peck, L. R., & Bell, S. H. (2014). The role of program quality in determining Head Start's impact on child development. OPRE Report #2014-10, Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.

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.

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(1), 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.

Santagata, R., Kersting, N., Givvin, K. B., & Stigler, J. W. (2011). Problem implementation as a lever for change: An experimental study of the effects of a professional development program on students' mathematics learning. Journal of Research on Educational Effectiveness, 4, 1–24.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). Lifetime effects: The High/Scope Perry Preschool study through age 40. Ypsilanti, MI: High/Scope Press.

Schweinhart, L. J., & Weikart, D. P. (1997). The High/Scope preschool curriculum comparison study through age 23. Early childhood research quarterly, 12(2), 117-143.

Sroufe, L. A., Coffino, B., & Carlson, E. A. (2010). Conceptualizing the role of early experience: Lessons from the Minnesota longitudinal study. Developmental Review, 30(1), 36-51.

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(5), 1435-1456.

U.S. Department of Health and Human Services, Administration for Children and Families (January 2010). Head Start Impact Study. Final Report. Washington, DC.

Walters, C. R. (2015). Inputs in the production of early childhood human capital: Evidence from Head Start. American Economic Journal: Applied Economics, 7(4), 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(5), 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(1), 122–154.

Yeh, S. S. (2010). The cost effectiveness of 22 approaches for raising student achievement.

Journal of Education Finance, 36(1), 38-75.

Online Appendix A: Additional Tables

Appendix 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 2-day/14 hour training, whereas all other TEEM courses were 2 hour duration.

Appendix 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

Appendix Table A3

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)

Appendix Table A4

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	
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	
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	
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

Panel B2: parent outcomes that fall under the construct responsive language and behavior support	
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

Appendix Table A5

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)

Note: 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.

Appendix Table A6

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
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

Continued on next page

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
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
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

Continued on next page

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.

Appendix Table A7

Total and per-pupil yearly costs of implementing the PALS intervention as planned and for two sites (detailed version of Table 5)

		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
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
c. 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

Continued on next page

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 year)		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.

Online Appendix B: Additional Information on Student Outcome Measures

A total of 17 separate student outcome measures were assessed at the beginning and end of each school year (results for each outcome are shown in Table 3). We tracked 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 highlighted above and align with the intended outcomes of the interventions. We used 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 (five non-cognitive outcomes including book reading engagement, shared enjoyment, enthusiasm / initiative, cooperation, and social engagement, and one cognitive outcome labeled language use);
- Student "tasks" (three non-cognitive outcomes used to measure self-regulation and executive functioning);
- Teacher and parent surveys (five non-cognitive outcomes used to assess social-emotional functioning, school liking, and school avoidance); and
- Written and oral exams (three cognitive outcomes used to assess language and literacy skills).

Observation ratings of child behavior. Six outcome measures were assessed during observations of child behaviors, which took place during free play and book reading sessions (described in the main body text). All observation scores were measured on a scale of one to five, with one indicating almost never and 5 indicating almost always. Child activities were video-taped for later coding by researchers (see main text for coding and rater training procedures). During video-taped free play sessions, children were rated on previously validated scales (Authors, year) that measure social engagement and cooperation. The social engagement scale had an intraclass correlation coefficient (ICC) of 0.65 and for cooperation, ICC = 0.71, indicating that most of the variation in our scales was between observations, rather than within an individual student observation. In the book reading sessions, students were 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 Appendix Table A4 and have demonstrated high reliability and validity in previous research (Authors, year).

Task completion. We used three special tasks to assess students' self-regulation and executive function before and after intervention. Self-regulation was measured through a gift delay-wrap task (Kochanska, Murray, & Harlan, 2000; Li-Grining, 2007), in which children were told that they would receive a present but that they could not turn and peek while the present was 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) were 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 second until first peek were strongly correlated (r = .76 at pretest; r = .78 at posttest) and were therefore standardized and averaged to create a total score (ICC = .94). A second test of self-regulation was the gift delay-bow task (Kochanska et al., 2000). The wrapped gift was placed in front of the child, who was told not to touch the gift until the examiner returned with a bow. The delay lasted two minutes. Strategy scores (on a scale of 1-5, e.g., 1 = removes toy; 5 = does not touch box) given for every 15 seconds of the task were averaged to create a total strategy score for the task. Scores for strategy and number of second until first touch, were strongly correlated (r = .62 at pretest; r = .63 at posttest), and were thus standardized and averaged to create a gift delay-bow total score. Finally, we measured changes in students' executive function using the bear/dragon task (Carlson, 2005; Garon et al., 2008). Children are told to follow the bear puppet's commands (e.g., touch your ear), but not the dragon puppet's commands. A pass/fail score was used in analyses based on recommendations from previous studies (5/6 correct dragon trials constituted passing the task; Carlson, 2005). For each of these three tasks, the interrater reliability was high, with ICC greater than 0.94 for each.

Teacher and parent surveys. Socio-emotional functioning and students' school liking and avoidance were measured using surveys of parents and teachers, resulting in a total of five non-cognitive outcomes. Both parents and teachers completed the Children's Behavior Questionnaire (CBQ; Rothbart, Ahadi, Hershey, & Fisher, 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 scale from one (extremely untrue of the child) to seven (extremely true of the child). The CBQ scales have demonstrated 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 found in the current study that for the parent-rated CBQ total score, internal consistency (Cronbach's α) was 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 resulted in two non-cognitive outcomes, the CBQ total parent score, and the CBQ total teacher score.

Parents and teachers also completed 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 found that each of these measures had adequate internal consistency for the parent-rated variables and high internal consistency for teacher-rated variables (Cronbach's α between 0.74 and 0.81 for parent ratings on pretests and posttests and between 0.80 and 0.93 for teacher-rated pre and posttest). Finally, teachers completed the School Liking and Avoidance Questionnaire (SLAQ; Ladd & Dinella, 2009), in which teachers used 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 times measuring school liking (e.g., enjoys most classroom activities). Internal consistency was α = 0.82 to 0.85 at pretest and α = 0.78 to 0.83 at posttest. because teacher and parent surveys were administered in person, we received a 100% response rate for all teachers and parents participating in the study.

Paper and pencil and oral exams. We used 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 assesses receptively understood language and the ability to verbally communicate with others. Last, the Test of Preschool Early Literacy (TOPEL; Lonigan, Wagner, Torgesen, & Rashotte, 2007) consists of three subtests: Print Knowledge (36 items), Phonological Awareness (27 items), and Definitional Vocabulary (35 items). As with all other student outcomes, assessments were taken before and after intervention each year and we include additional information in the online appendix.

Due to high correlations among the language measures from the students' cognitive tests, we formed a composite "language skills" factor consisting of the EOWPVT, both PLS subscales, and the Definitional Vocabulary subscale of the TOPEL. We found high internal consistency of this measure at both pretest (Cronbach's α = 0.93) and posttest (α = 0.92).