Early Childhood Research Quarterly 59 (2022) 56–73

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

journal homepage: www.elsevier.com/locate/ecresq

Program-wide implementation of the Pyramid Model: Supporting fidelity at the program and classroom levels

Mary Louise Hemmeter[a,*], Erin Barton[a], Lise Fox[b], Christopher Vatland[b], Gary Henry[c], Lam Pham[d], Kymberly Horth[a], Abby Taylor[a], Denise Perez Binder[b], Meghan von der Embse[b], Myrna Veguilla[b]

[a] Vanderbilt University, Department of Special Education, Peabody College, Box 228 GPC, Nashville, TN 37203
[b] University of South Florida, Department of Child and Family Studies, 13301 Bruce B Downs Blvd, MHC 2113A, Tampa, FL 33612
[c] University of Delaware, School of Education and Biden School of Public Policy and Administration, 106 Alison Hall West, University of Delaware, Newark, DE 19716
[d] North Carolina State University, Department of Educational Leadership, Policy, and Human Development, Poe Hall 300P, 2310 Stinson Drive, Raleigh, NC 27695

ARTICLE INFO

Article history:
Received 20 September 2020
Revised 2 October 2021
Accepted 14 October 2021
Available online 28 November 2021

Keywords:
Social-Emotional
Implementation
Pyramid Model
Challenging Behavior
Fidelity

ABSTRACT

Many early childhood programs are not prepared to meet the needs of children who have significant social, emotional, and behavioral challenges. *Program-Wide Supports for Pyramid Model Implementation* (PWS-PMI) provides a systematic approach to supporting early childhood programs using Pyramid Model practices and enhancing children's social-emotional outcomes that is grounded in implementation science. We designed the current study to test the PWS-PMI intervention and examine its feasibility of implementation in community-based early childhood programs serving children from low-income environments. In this study, we found programs increased their implementation of PWS-PM and improved classroom practices after only one year of support. Our findings indicate a program-wide approach is effective, although more time and support will be necessary to sustain high fidelity implementation and produce robust effects on children.

1. Introduction

Many early childhood programs are not prepared to meet the needs of children with significant social, emotional, and behavioral challenges. Early childhood teachers report challenging behavior to be a significant professional development need (Reinke et al., 2011; Snell et al., 2012). High rates of suspension and expulsion from early childhood programs highlight the limited capacity of programs to meet the needs of children with challenging behavior (Meek & Gilliam, 2016). Alarmingly, these data also show that exclusionary discipline is used disproportionately with Black boys who are suspended and expelled at higher rates than their peers (U.S. Department of Education Office of Civil Rights, 2014; 2016). Evidence that exclusionary discipline actions are being applied disproportionally raises concerns related to equity and bias. National concerns about the use of suspension and expulsion in early childhood programs and the evidence of bias in their use led to a national policy statement that calls for states and programs to implement a variety of recommendations related to creating positive climates and building workforce capacity to promote children's social emotional and behavioral health and appropriately intervene when there is challenging behavior (U.S. Department of Health and Human Services & U.S. Department of Education, 2014).

While there are effective interventions for addressing the social, emotional, and behavioral needs of young children (Barton et al., 2014; Domitrovich et al., 2007; Hemmeter et al., 2021a; Webster-Stratton et al., 2008), there is a gap between research on the practices and implementation of the practices in community-based programs (e.g., Dunst & Trivette, 2009; Greenberg et al., 2017; Metz & Bartley, 2012; Odom, 2009). Researchers have noted that program-level implementation supports are critical to implementation fidelity by practitioners and the sustained use of evidence-based practices (Bierman & Motamedi, 2015; Durlak, 2015; Halle et al., 2013, Oberle et al., 2016; Webster-Stratton & McCoy, 2015). In the field of social emotional learning in schools, researchers also have championed the notion that a whole school approach is needed to create a supportive context where effective implementation of social emotional interventions can be used and sustained (Durlak, 2015; Greenberg et al., 2003; Oberle et al., 2016; Osher et al., 2016). Implementation challenges that may compromise whether an intervention is used within a

* Corresponding author.
E-mail address: ml.hemmeter@vanderbilt.edu (M.L. Hemmeter).

https://doi.org/10.1016/j.ecresq.2021.10.003
0885-2006/

program are often related to issues of buy-in, organizational capacity, staff training, data monitoring, and dedicated resources needed for sustainability (Domitrovich et al., 2012; Durlak, 2015; Wenz-Gross & Upshur, 2012). Much of the work on social emotional interventions in early childhood settings is focused on the individual and classroom teacher level rather than evaluation of program-wide implementation of social emotional programs. As a result, there is a lack of information about effective strategies that might be used to build the capacity of an early childhood program to implement and sustain an intervention program or practices by all teachers over time.

2. Pyramid Model for Promoting Social Emotional Competence of Young Children

One approach to addressing the social, emotional, and behavioral needs of young children is the *Pyramid Model for Promoting the Social Emotional Competence of Young Children* (Hemmeter et al., 2021b). The Pyramid Model is a framework for organizing evidence-based practices that include universal social-emotional promotion practices for all children, practices for children who need targeted social-emotional supports, and individualized behavior support practices for children with significant social skill deficits or persistent challenging behavior. Pyramid Model practices are based on research on effective instruction (Burchinal et al., 2010; National Research Council, 2001), supporting engagement (Chien et al., 2010; Conroy et al., 2008; Covington-Smith et al., 2011), promoting social skills and emotional competencies (Domitrovich et al., 2012), and using assessment-based behavior support plans for children with severe and persistent behavior challenges (Blair et al., 2011; Dunlap et al., 2018).

Teachers in early childhood settings, without training and support, implement Pyramid Model practices at about 40% fidelity (Hemmeter et al., 2013). This level of implementation is not adequate for affecting changes in children's social, emotional, and behavioral outcomes (Hemmeter et al., 2016, 2021a). Two randomized trials examined the effects of professional development (PD) on teachers' use of Pyramid Model practices and children's social skills and problem behavior (Hemmeter et al., 2016; Hemmeter et al., 2021a). The studies demonstrated the effectiveness of the PD intervention on teachers' implementation of Pyramid Model practices and suggested implementation fidelity is related to children's social-emotional outcomes.

3. Need for Program-Wide Supports

A challenge to the implementation of the Pyramid Model in early childhood programs is the delivery of sufficient and ongoing professional development to achieve a level of implementation like what has been achieved in research on the approach. While research has documented the effects of a systematic professional development (PD) intervention on teachers' use of the Pyramid Model, existing studies have been conducted only with individual teachers. Although the PD intervention was successful with 16 weeks of individualized coaching, many early childhood programs do not have the capacity to provide that level of support to each teacher. It is probable that within programs, some teachers would need more supports and some less, and the level of support might vary depending on other implementation supports (e.g., a program-wide commitment to behavior support, family engagement in program-wide efforts).

The critical importance of providing implementation supports has been noted by researchers examining factors that influence the relationship between implementation and outcomes of evidence-based interventions in applied settings (Domitrovich & Greenberg, 2000; Durlak, 2015; McIntosh et al., 2013). Examples of implementation factors include professional development features that promote fidelity by practitioners and the factors (e.g., shared vision, decision-making, leadership, administrative support) related to the operation of program that hosts the intervention (Domitrovich et al., 2008; Durlak & Dupre, 2008).

The aim of this study was to design and evaluate an intervention to provide program-wide supports for implementing the Pyramid Model. In program-wide implementation of the Pyramid Model, a leadership team guides the implementation of critical elements needed to promote implementation. Critical elements include establishing staff buy-in, engaging with families, establishing program-wide expectations, providing ongoing professional development and classroom coaching, establishing procedures for responding to the needs of children with challenging behavior, and using data to monitor implementation and outcomes (Fox et al., 2013).

Program-wide supports are designed to create a context where support is systemic, sustained, and likely to have lasting impacts. Further, we need an approach that is feasible, effective, and efficient, with particular attention to the limited time and resources of early childhood programs. While evidence supports the effects of the Pyramid Model (Hemmeter et al., 2016; 2021a), there are no studies examining a program-level intervention focused on increasing a set of organizational practices that are likely needed to support implementation of the Pyramid Model throughout program. In this study, we address this by delivering an intervention known as Program-Wide Supports for Pyramid Model Implementation (PWS-PMI).

We designed the current study to test the PWS-PMI intervention and examine its feasibility of implementation in early childhood programs serving children from families with low incomes. There are several layers to our study. First, the PWS-PMI intervention is a cohesive set of processes and practices for programs to use in delivering program-wide supports. Programs were guided by an external coach (i.e., research staff) who provided support to the leadership team as they developed and implemented program-wide supports. Following the conceptual framework for examining variation in program effects by Weiss et al. (2014), we define intervention fidelity as the extent to which the external coach adhered to the delivery of support for program staff in delivering the PWS-PMI. Treatment fidelity, its own distinct type of fidelity, is defined as the extent to which program personnel take up the intervention and implement the processes and practices (Weiss et al., 2014). Treatment contrast (Hulleman & Cordray, 2009) is the difference between treatment fidelity in the treatment and control groups. A goal of this study was to examine the effects of the intervention on classroom implementation of the Pyramid Model. Also, we examined the effect of the intervention on overall classroom quality. A final goal was to examine the effects on program-level behavior variables and children's social skills and challenging behaviors. The research questions were: (a) Does the PWS-PMI intervention affect treatment fidelity or the take-up of program-wide supports for implementing the Pyramid Model, program-wide implementation of positive behavioral supports, or behavior events? (b) Does the PWS-PMI intervention increase classroom implementation of Pyramid Model practices and/or overall classroom quality? and (c) Does the PWS-PMI increase children's social skills and decrease children's problem behavior?

4. Method

4.1. Design

We implemented a cluster randomized controlled trial (RCT) with random assignment to conditions occurring at the program level. We defined the study population as early childhood pro-

Table 1
Descriptive Program, Child, and Teacher Characteristics at Baseline and Post-Intervention.

	Baseline				Post-Intervention			
	Control	Treated	Difference	p-value	Control	Treated	Difference	p-value
Program Characteristics								
Tuition Fee	0.75	0.71	-0.04	0.89	0.88	0.71	-0.16	0.47
Head Start/EHS	0.12	0.57	0.45	0.08	0.12	0.57	0.45	0.08
External Support with Behavior	0.62	0.43	-0.20	0.48	0.88	0.71	-0.16	0.47
Child Characteristics								
Average Age (Years)	3.88	3.95	0.07	0.53	4.39	4.27	-0.12	0.48
Proportion Girls	0.44	0.49	0.05	0.29	0.44	0.50	0.06	0.43
Proportion Black	0.27	0.57	0.30	0.09	0.27	0.57	0.30	0.09
Proportion Hispanic	0.14	0.04	-0.09	0.30	0.14	0.04	-0.09	0.30
Proportion Asian	0.02	0.02	0.01	0.75	0.02	0.02	0.01	0.75
Proportion White	0.40	0.19	-0.21	0.22	0.40	0.19	-0.21	0.22
Proportion Multiple Race	0.09	0.09	0.01	0.91	0.09	0.09	0.01	0.91
Proportion Other Race	0.00	0.09	0.08	0.30	0.00	0.09	0.08	0.30
Proportion Unknown Race	0.08	0.00	-0.08	0.37	0.00	0.00	-0.00	0.37
Proportion IEP/IFSP	0.02	0.02	0.00	0.57	0.03	0.01	-0.01	0.41
Proportion Dual Language Learner	0.17	0.14	-0.03	0.78	0.10	0.14	0.04	0.77
Proportion Child Care Subsidy	0.44	0.57	0.14	0.50	0.44	0.55	0.11	0.55
Teacher Characteristics								
Proportion Female Teachers	1.00	1.00	0.00	.	1.00	1.00	0.00	.
Proportion Black Teachers	0.41	0.76	0.35	0.16	0.49	0.75	0.26	0.30
Proportion White Teachers	0.46	0.29	-0.17	0.51	0.42	0.29	-0.13	0.62
Proportion Teachers with BA or Above	0.25	0.35	0.10	0.69	0.46	0.31	-0.15	0.54
Proportion Teachers with ECE Degree	0.77	0.60	-0.17	0.40	0.76	0.52	-0.23	0.29
Proportion Teachers with Pre-K License	0.44	0.42	-0.03	0.89	0.40	0.47	0.07	0.76
Average Teacher Experience (Months)	135.38	136.89	1.51	0.97	149.42	149.76	0.33	0.99
Average Months of Experience in Position	39.90	46.65	6.75	0.77	48.60	59.17	10.57	0.67

EHS: Early Head Start; IEP: Individual Education Plan; IFSP: Individualized Family Support Plan; ECE: Early Childhood Education.
Significance comparisons are made using t-tests between treated and control groups. The social skills scale and problem behaviors scale have been standardized to have a mean zero and unit variance.
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

grams that serve children ages 2-5 and recruited programs in two states. Before assigning the programs to conditions, pairs of programs were placed into blocks based on state, whether the program was a Head Start center, and the number of classrooms. This process created eight blocks of two programs each. Within each block, one program was randomly assigned to receive the PWS-PMI intervention and the other program served as a control.

4.2. Participants

Sixteen programs were randomly assigned. Eight were assigned to the PWS-PMI intervention condition, and eight to the business-as-usual control condition (BAU).

4.2.1. Programs

The 16 programs were in two southern states and included Head Start, for-profit private, church-affiliated, and nonprofit community programs (Table 1). We intentionally recruited a range of programs in terms of size, race and ethnicity of children, and enrollment of children with disabilities. A program in the intervention condition closed during the intervention; thus, we were unable to collect post-intervention data. Of the 15 programs that completed the study, five were Head Start, 11 were tuition/fee-based childcare, and six were state-funded Pre-K programs. Some community-based child care programs included both public Pre-K and Head Start classrooms. Ten of the programs reported using some type of social-emotional learning program. The groups did not vary in terms of how much support they received from professionals (e.g., mental health consultants, behavior consultants) outside of the program (i.e., external behavior support; 62% in control and 43% in treatment) prior to the study. There were no differences between groups in terms of exposure to the Pyramid Model prior to the study.

We recruited administrators, classroom coaches, and teaching teams from participating programs to serve on the program's leadership team. Leadership teams comprised two to six administrators, classroom coaches, behavior support specialists, or teachers. In some programs, the same person served in multiple roles (e.g., one person was a classroom coach and behavior support specialist). We recruited 81 leadership team members (38 intervention; 29 control; 14 leadership team members left their programs and thus were withdrawn from the study).

4.2.2. Classrooms

In addition to program leaders, we recruited teachers from the classrooms of children ages 2-5 years at each program (Table 1). We recruited 56 teachers (33 intervention; 23 control) at the beginning of the study, but the number of teachers who participated in data collection varied as the study progressed because of staffing changes and teacher turnover. These changes included, but were not limited to, one program closing, teachers moving from preschool to infant classrooms which did not qualify for data collection, and teachers resigning from the program. With these staffing changes, we obtained complete data on 40 teachers for final analysis (17 in intervention programs and 23 in control programs).

4.2.3. Child Participants

To determine the effects of PWS-PMI on children's social skills and challenging behavior, data were collected on all children ages 3-5 years in the participating classrooms ($n = 589$; intervention $n = 331$; control $n = 258$). While there were 2-year-olds in some of these classrooms, we did not collect data on children under 3. As with teachers, the number of children varied across time, because some children left classrooms or programs participating in the study and new children moved in throughout the year. Given these changes, our final analytic sample included 335 children (197 in treatment programs, 138 in control programs) with complete data collected before and after the implementation of PWS-PMI.

Table 2
Intervention Fidelity.

	Monthly Meetings	Weekly Visits
Average coaching log indicators implemented (range)	81% (52-100)	71% (36-99)
Average number of strategies used per session (range)	6.1 (1-16)	4.8 (1-19)
Most Common Activities	Attend LT meeting Respond to questions Review implementation plan	Respond to questions, Observe teacher(s) without coach Observe teacher(s) with coach

LT: Leadership Team.

4.3. PWS-PMI Intervention

In this section, we describe the procedures used with programs in the intervention and control conditions. Programs assigned to the control condition received no supports during the study year but received training in the year following data collection.

4.3.1. Intervention Fidelity in the Treatment Condition

Programs assigned to the PWS-PMI condition received the following supports: (a) leadership team, coach, teacher, and behavior support training; and (b) ongoing support from an external coach. We describe fidelity with which the intervention was delivered below.

Training Workshops. We conducted five in-person workshops for each intervention site: (a) a 2-day leadership team workshop; (b) a 2-day Pyramid Model practices workshop for teaching staff; (c) a 1-day workshop for behavior support personnel (Dunlap et al., 2013); (d) a 1-day workshop for classroom coaches on practice-based coaching (PBC; Snyder et al., 2015); and (e) a 1-day training for classroom coaches on the Teaching Pyramid Observation Tool (TPOT; Hemmeter et al., 2014). While a leadership team was formed and roles were identified for the control group programs, all workshops for the control group were conducted after study data were collected.

At the conclusion of each training event, anonymous participant evaluations were collected via paper-pencil forms. Participants answered 10 questions for each event on a Likert-type scale (*strongly disagree*=1 through *strongly agree*=4). Questions addressed the following topics: (a) the trainer's presentation of content; (b) appropriateness or applicability of the content; and (c) feasibility of implementation. Mean scores across training events indicate participants were satisfied with the trainings ($M = 3.68$ for intervention group training events, and $M = 3.5$ for control group training events) (See Appendix Table A1).

External Coaching. Following the leadership team training, programs received individualized coaching weekly to guide their program-wide implementation of the Pyramid Model. The external coach made weekly visits and attended monthly leadership team meetings to assist with activities including, but not limited to, planning leadership team meetings, supporting the classroom coach, supporting the behavior support personnel, and assisting with family events. The external coach attended an average of nine monthly leadership team meetings (*range* = 6-10) and made an average of 25 other weekly visits (*range* = 14-36). A summary of the frequency and type of supports provided to each program is shown in Appendix Table A2.

Control Condition. Programs in the control condition received monthly check-in emails or phone calls related to data collection and study scheduling. Research staff provided no support related to Pyramid Model implementation during the study year. Control programs were invited to participate in the training workshops during the summer after data collection ended.

Intervention Fidelity. Three types of data were collected to measure intervention fidelity or the delivery of the intervention to program personnel, which is separate and distinct from fidelity of implementation of the treatment by the programs. Fidelity measures included information on (a) duration of and content covered in workshops, (b) duration of and adherence to an external coaching protocol, and (c) types of strategies used during external coaching. Training events were scored for fidelity related to: (a) duration, (b) content, and (c) training activities and strategies. Fidelity for each training event was above 90% with the exception of one training for the teaching staff (scores for each event are provided in Appendix Table A1). External coaches completed logs for each session. The logs included (a) indicators that measured fidelity to the coaching protocol, (b) activities they engaged in during their external coaching visits (e.g., observe with classroom coach, attend a family event), and (c) information about duration of visits. Coaching log data (i.e., intervention fidelity) are summarized in Table 2.

4.4. Measures

Data were collected at the program, classroom, and child level in experimental and control classrooms. Unless noted, all data were collected at baseline and post-intervention.

4.4.1. Program Measures

Supporting Program-Wide Implementation Fidelity Inventory (SPIFI). We developed the SPIFI to measure treatment fidelity. The SPIFI includes nine items that represent key components of PWS-PMI with a total of 82 indicators. The nine items include (a) leadership team composition, (b) leadership team activities, (c) staff buy-in, (d) development and implementation of program-wide expectations, (e) procedures for developing behavior support plans, (f) staff support plan, (g) family engagement around the program-wide plan, (h) family engagement related to supports for individual children, and (i) data-based decision making. Data collection is conducted using multiple sources of evidence from program, including staff and family interviews, brief observations, and a document review. Each item is scored from zero to seven based on the number of yes/no responses for each indicator. The SPIFI has high reliability ($\alpha > 0.85$) and good convergent validity with the Preschool-wide Evaluation Tool (PreSET; Steed et al., 2012) (authors' analysis, available on request). We operationalize scoring the SPIFI outcome indicators in two different ways to show that outcomes are not sensitive to how the instrument is scored. The two scoring methods we use are (a) the sum of all 82 individual indicators on the instrument (a score ranging from 0 to 82) or (b) the sum of the scores averaged across the nine SPIFI items (a score ranging from 0 to 63). Both scores are standardized to have a mean of zero and standard deviation of one.

PreSET. The PreSET (Steed et al., 2012) is an instrument used to evaluate the fidelity of the universal tier of positive behavioral interventions and supports in an early childhood setting. PreSET data are collected by an outside observer who conducts observations and interviews to assess implementation. The range of scale scores for PreSET is 0-100, and scores indicate the percentage of indicators observed within each subscale. These scale scores are standardized to have a mean of zero and a standard deviation of one. Analyses have been conducted on inter-rater reliability,

test-retest reliability, and construct validity of the PreSET (Steed & Webb, 2013). Inter-rater reliability yielded an agreement of 95% and an overall *k* value of .80. The PreSET showed good internal consistency, with an alpha of .91, and strong item-subscale correlations with a mean of .56 and a median of .58.

Program Demographic Questionnaire. The program demographic questionnaire addressed program demographics and leadership team composition and experience. Administrators completed the questionnaire with updated enrollment information during all data collection periods, and data from baseline and post-intervention were used in the analyses.

Discipline Questionnaire. The program discipline questionnaire asked program administrators to document the following: (a) Were any child(ren) dismissed from program due to challenging behavior or given the child's behavior support needs; (b) Were any child(ren) transferred to another program that would be better able to meet child's behavior support needs; (c) Were any child(ren) asked to stay home for one day or more due to challenging behavior with a return to the program; and (d) Were any child(ren) sent home for remainder of day due to challenging behavior. If the answer to any of these questions was "yes", then administrators were asked to document how many children for each question. Data collected during baseline and post-intervention were used in analyses. We conceptualized discipline as the total number of disciplinary incidents that occurred, but we discuss below how our conclusions do not change when we operationalize discipline in different ways (e.g., if any disciplinary action occurred).

4.4.2. Classroom Measures

Teaching Pyramid Observation Tool (TPOT). The TPOT (Hemmeter, Fox, & Snyder, 2014) is a tool for measuring teachers' implementation of the Pyramid Model practices in the classroom. The TPOT has three subscales: (a) key teaching practices (14 items), (b) red flags (17 items), and (c) effective strategies for responding to challenging behavior (3 items). The TPOT is completed following a 2 hour observation in the classroom and an interview with the teacher. Research has examined the psychometric integrity of the TPOT and demonstrated that it is sensitive to changes in teachers' practices related to the Pyramid Model (Snyder et al., 2013). A generalizability theory study (G-study; Shavelson & Webb, 1991) showed minimal error variance (5%) attributed to occasions and raters, and the generalizability coefficient was .97. TPOT data were standardized to have a mean of zero and a standard deviation of one.

Classroom Assessment Scoring System (CLASS). The CLASS (Pianta et al., 2008) was used to measure the extent to which implementation of the Pyramid Model impacts classroom quality. Ten items are organized under three domains: (a) emotional support ($v = 4$), (b) classroom organization ($v = 3$), and (c) instructional support ($v = 3$). Scores for dimensions and domains on the CLASS range from 1 (*low*) to 7 (*high*). The instrument has been demonstrated to have high inter-rater score reliability across dimensions ranging from 78.8 to 96.9. Internal consistency score reliability estimates range from .79 to .91 in preschool classrooms. Confirmatory factor analyses, using data from five samples, support the theoretical structure of the measure. Structure coefficients ranged from .69 to .96. Goodness-of-fit indices ranged from .89 to .97 across samples, and comparative fit indices ranged from .93 to .96. The composite score for each of the three CLASS domains are standardized to have a mean of zero and a standard deviation of one.

Classroom Demographic Measures. The classroom demographic questionnaire was completed by consented teachers in classrooms of children ages 2 to 5 years during baseline and post-intervention. Teachers answered questions about themselves (i.e., gender, race/ethnicity, degrees, training, certification) and other adults in their classroom (i.e., number of adults and roles, how much time they spend in the classroom daily).

4.4.3. Child Measures

Social Skills Improvement System Rating Scales (SSIS). Teachers completed the SSIS (Gresham & Elliot, 2008) for each child in their classroom during baseline and post-intervention as a measure of children's social skills and problem behavior. Higher scores for social skills are positive whereas higher problem behavior scores are concerning and suggest a need for remediation. Scale scores and percentile ranks are provided for both social skills and problem behaviors. We also standardize the scale scores to have a mean of zero and unit standard deviation of one. Median scale score reliabilities of the scales are in the mid- to upper .90s for every age group on each form. Test-retest score reliability ranges from .68 to .86 with a mean adjusted coefficient of .81 (Gresham & Elliot, 2008).

4.5. Data Collection Procedures

We had measures at the program, classroom, teacher, and child levels. All measures were administered in both the intervention and control groups, and observational measures were conducted on different days. The baseline (i.e., pre-intervention) measures were collected between May and September of 2017; post-intervention measures were collected between April and June of 2018. All measurements were from live classroom observations by data collectors who were masked to the treatment status; data collectors were recalibrated prior to each wave.

4.5.1. Data Collector Training

Data collectors were graduate students in Early Childhood Special Education, Child Studies, or a related field. They were naïve to the condition to which the program was assigned. They were trained to a minimum of .80 interrater agreement on measures prior to data collection. During data collection, a second observer conducted simultaneous reliability observations on a minimum of 25% of observations and assessments. Data collector training was conducted to ensure reliability and consistency in the use of study measures. A data collection manual was created to train data collectors. Training occurred in two stages. First, the project coordinator (PC) conducted a training for the lead data collectors from both sites. Training on all tools involved the following steps: (a) lead data collectors read the manual, (b) PC provided an overview of the administration procedures, (c) PC and lead data collectors watched videos, scored the instruments, and (d) PC and lead data collectors observed classrooms and scored instruments until they achieved 80% agreement. Once lead data collectors reached criterion, they used the same procedures with all data collectors until 80% interrater agreement was reached.

4.5.2. Inter-observer Agreement

During the study, inter-observer agreement (IOA) data were collected on all measures, all teachers, and all observers for at least 25% of observations. Data were double entered to ensure the accuracy of coding. For the SPIFI, CLASS, TPOT, and PreSET, we tested IOA using intraclass correlation coefficients (ICCs) from multilevel models where rater evaluations were nested within programs. The ICC is the proportion of observed variance that occurs between programs; therefore, the ICC is high when the variation between programs is large relative to variation between evaluators observing the same program. We found substantial reliability both when the SPIFI was scored as either the sum of all items (ICC = 86%) or the sum of all indicators (ICC = 96%). We also found high ICCs for

the three CLASS domains: emotional support (ICC = 94%), classroom organization (ICC = 94%), and instructional support (92%). For the overall average scores of TPOT and PreSET, we found ICCs of 93% and 99%, respectively. These high ICC values provide evidence of interobserver agreement across all measures.

4.6. Analytic Sample

The analytic sample includes 15 programs, 40 teachers' classrooms and 335 children with complete data at both the baseline and post-intervention periods. Since one program withdrew from the study, the analytic sample is somewhat reduced from the full sample that was originally part of the random assignment. Specifically, 82 children and 10 teachers from PWS-PMI programs attrited from the sample relative to 78 students and 8 teachers in control programs. At the child level, the numbers of attrited children correspond with an overall attrition rate of 32%, an attrition rate of 29% in treated programs, and an attrition rate of 36% in control programs. At the classroom level, the numbers of attrited classrooms correspond with an overall attrition rate of 31%, an attrition rate of 30% in treated programs, and an attrition rate of 32% in control programs. Following the cautious thresholds of $\alpha_{treatment} = .27$ and $\alpha_{control} = .22$ from the What Works Clearinghouse, these attrition levels correspond with a potential bias of approximately .02 standard deviation units (i.e., low attrition) (WWC, 2013, 2014). Additionally, in Appendix Table A3, we show evidence that program characteristics are similar in the baseline and post-intervention periods, providing evidence that results are not driven by sample attrition.

Appendix Table A3 shows the descriptive means for all study participants during baseline and post-intervention. About one-third of the participating programs were either Head Start or Early Head Start sites. Participating programs primarily served Black children (40%) and White children (30%). Most of the programs charged tuition (70%-80%) but about 50% of children received a childcare subsidy. Teachers in the sample were all female and either Black (about 60%) or White (about 35%), with very few teachers identifying in other race categories. On average, teachers had about 12 years of teaching experience with about four years at their current position. There were no significant changes in program, teacher, or child covariates based on measurements taken at the baseline across the two time points, except the naturally occurring increase in the average student age. These results were not driven by changes in our sample across time (i.e., sample attrition).

Before formally estimating the effects of PWS-PMI, we conducted descriptive *t*-tests comparing child, teacher, and program characteristics between the intervention and control groups in both the baseline and post-intervention periods. These *t*-tests at baseline allowed us to examine whether random assignment successfully resulted in balanced or observably similar intervention and control groups. As shown in the first four columns of Table 1, none of the demographic characteristics differed at statistically significant levels at baseline, suggesting that our random assignment produced baseline balance of intervention and control groups on observed covariates. Conducting *t*-tests at baseline also allows us to better understand whether certain program characteristics should be included as covariates, to adjust for initial imbalances. Although no variables were significantly different at baseline, there are some characteristics that differed somewhat in magnitude between intervention and control groups. Compared to control programs, programs assigned to PWS-PMI intervention were less likely to charge tuition (71%), more likely to be a Head Start program (57%), and less likely to already be implementing a behavior support program (43%). Moreover, intervention programs tended to serve a larger proportion of Black children (57%) and a smaller proportion of White children (19%). Teachers at treated programs were more likely to be Black (76%) and have a bachelor's degree or higher (35%). To correct for these imbalances in baseline characteristics that could influence our estimates of effects, we adjusted for these covariates in our analytic models (described below). We also show, in Appendix Tables A6 and A7, that our results are robust to the inclusion of different covariates.

Additionally, the descriptive *t*-tests in the post-intervention period allows us to examine whether characteristics of the treated and control programs changed in ways that could potentially bias our results. Table 1 shows that no program, teacher, or child characteristic differs at statistically significant levels in the post-intervention period, providing evidence that any potential differences in outcomes are not driven by observable changes in either the programs themselves or in characteristics of teachers and children in these programs.

4.7. Analytic Models

To formally estimate the effects of PWS-PMI, we use ordinary least squares regression (OLS) to examine program-level outcomes and use a series of hierarchical linear models (HLM) to examine classroom and student-level outcomes. For program level outcomes (PreSET, SPIFI, and disciplinary actions), we use OLS with covariate-adjustment instead of HLM because the program-level outcomes are not nested within higher level units. To conduct the analysis, we use a model build-up approach that adds an increasing number of covariates, building up to Equation 1 below (Hox et al., 2010). Equation 1 models each of the three program-level outcomes y for each program k at time t. First, we include only an indicator for the PWS-PMI intervention. This model with only the treatment indicator ($Treat_{kt}$) effectively estimates the unconditional mean outcome difference between intervention and control programs. Then, we add a baseline measure of the dependent variable to help control for any initial differences between intervention and control programs prior to intervention that are not controlled by random assignment. Finally, to improve statistical power to detect effects, we include a set of program-level characteristics that differed in magnitude between intervention and control groups, all measured before the PWS-PMI intervention began. These covariates were selected based on the previously described qualitative but not statistically significant differences between treated and control programs at baseline and are the same for every model. The covariates include whether the program charges tuition, program type (i.e., whether the program is a Head Start/Early Head Start site or a community program), whether the program is already implementing an external behavior support system, the proportions of Black and White children (with other race as the reference category), the proportion Black teachers (with non-Black teachers as the reference category), and the proportion of teachers with a bachelor's degree or above. In Appendix Tables A6 and A7, we show that our results are robust to the inclusion of different covariates. In addition to these covariates, the model includes a block fixed effect (b_k) to reflect the blocked randomization design. Finally, Equation 1 includes a random error term at the program level (e_{kt}).

$$\begin{aligned} y_{kt} = {} & \gamma_0 + \gamma_1 Treat_{kt} + \gamma_2 y_{kt-1} + \gamma_3 Tuition_{kt-1} + \gamma_4 HeadStart_{kt-1} \\ & + \gamma_5 ExternalSupport_{kt-1} + \gamma_6 BlackStudents_{kt-1} \\ & + \gamma_7 WhiteStudents_{kt-1} + \gamma_8 BlackTeachers_{kt-1} \\ & + \gamma_9 BAAbove_{kt-1} + b_k + e_{kt} \end{aligned} \quad (1)$$

For the classroom level outcomes (TPOT and CLASS), we use two-level HLM models with classrooms j nested within programs k at time t. Following the same logic as the models for program-level outcomes, we use a model-build up approach where we begin with a null model that does not have any covariates. Then, we

add the treatment indicator and finally add the same set covariates as in Equation 1 above. The reduced form two-level model is shown in Equation 2 below. Equation 2 also includes a block fixed effect (b_k), a fixed effect for programs (u_{0kt}), and a random error term (e_{jkt}) at the classroom level.

$$\begin{aligned} y_{jkt} = {} & \gamma_{00} + \gamma_{01} Treat_{kt} + \gamma_{10} y_{jkt-1} + \gamma_{02} Tuition_{kt-1} \\ & + \gamma_{03} HeadStart_{kt-1} + \gamma_{04} ExternalSupport_{kt-1} \\ & + \gamma_{05} BlackStudents_{kt-1} + \gamma_{06} WhiteStudents_{kt-1} \\ & + \gamma_{07} BlackTeachers_{kt-1} + \gamma_{08} BAAbove_{kt-1} + b_k + u_{0kt} + e_{jkt} \end{aligned} \tag{2}$$

For child-level outcomes (i.e., the social skills and problem behavior measures from SSIS), we use HLM models with children nested within classrooms nested within programs. These hierarchical models are especially important in our context because the PWS-PMI intervention occurs at the program level. Since children in the same classrooms (and programs) are likely exposed to similar influences external to the PWS-PMI intervention, failure to account for the non-independence between children could downwardly bias the standard-errors of our estimates and increase the "false positive" errors in hypothesis testing (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). To model children (level 1) nested within classrooms (level 2) nested within programs (level 3), we begin with a null model:

Level 1: $y_{ijkt} = \pi_{0jkt} + e_{ijkt}$
Level 2: $\pi_{0jkt} = \beta_{00kt} + r_{0jkt}$
Level 3: $\beta_{00kt} = \gamma_{000} + u_{00kt}$

$$\text{Reduced Form}: y_{ijkt} = \gamma_{000} + u_{00kt} + r_{0jkt} + e_{ijkt} \tag{3}$$

In Equation 3, we model y (either social skills or problem behavior) for child i in classroom j and program k at time t. No predictors are included in the null model, but random intercepts for classrooms (r_{0jkt}) and programs (u_{00kt}) allow us to calculate intraclass correlations (ICCs) at both levels to examine how much of the variability in children's social skills and problem behavior SSIS scores exists at level 2 and level 3. With the inclusion of all predictors, the reduced form three-level model is:

$$\begin{aligned} y_{ijkt} = {} & \gamma_{000} + \gamma_{001} Treat_{kt} + \gamma_{100} y_{ijkt-1} + \gamma_{002} Tuition_{kt-1} \\ & + \gamma_{003} HeadStart_{kt-1} + \gamma_{004} ExternalSupport_{kt-1} \\ & + \gamma_{005} BlackStudents_{kt-1} + \gamma_{006} WhiteStudents_{kt-1} \\ & + \gamma_{007} BlackTeachers_{kt-1} + \gamma_{008} BAAbove_{kt-1} + b_k \\ & + u_{00kt} + r_{0jkt} + e_{ijkt} \end{aligned} \tag{4}$$

Since program randomization occurs within blocks, all models include a block fixed effect. The block fixed effect essentially compares the intervention program with the control program in each block and averages the differences across the sample. Note that the block fixed effect coupled with the small number of programs means that some program characteristics used as covariates do not vary within block and drop out of the model. Thus, at the program level, we report results from models that include only the baseline measure of the independent variable as a covariate. All data analyses were performed using Stata 15 (*StataCorp, 2015*).

5. Results

5.1. Outcome Measures

Table 3 shows program-level results for the standardized treatment fidelity measures (SPIFI, PreSET) to provide an estimate of the treatment contrast and the total reported number of student disciplinary actions. The effects of PWS-PMI on the standardized SPIFI total indicator and total item scores are 1.63 and 1.44 standard deviation units, respectively. These results show that PWS-PMI has a sizeable effect on program take-up of the PWS-PMI. The effect of PWS-PMI on the standardized PreSET score (1.28 standard deviation units) is also positive and statistically significant. Effect sizes (Hedges's g) for the SPIFI total indicator, SPIFI total item, PreSET are 2.26, 1.60, and 1.24, respectively. The parallel results on PreSET and SPIFI provide confidence that our findings are not attributable to the measure of implementation developed specifically for this intervention (i.e., SPIFI). Finally, we turn to the effect of PWS-PMI on disciplinary actions measured throughout the intervention, presented in Table 3. The point estimate on the number of disciplinary events is -0.08 but not statistically significant. While the estimated coefficient on the reduction of disciplinary actions is sizeable, the lack of statistical significance may be attributable to the statistical power to detect effects at the program level. As noted above, we constructed several alternative measures of disciplinary outcomes at the program level, all of which led to similar conclusions and are available upon request.

Table 4 shows the effect of PWS-PMI on the proximal outcomes measured by TPOT and CLASS. For brevity, we do not show results from the null model and model containing only the treatment indicator. However, for each classroom-level outcome, the null models show a sizeable amount of the variability occurs at the program level (ICC up to 16%) and support use of two-level models. Column 1 of Table 4 shows that when controlling only for the TPOT outcome at baseline, programs that receive the PWS-PMI intervention have TPOT scores that are 0.74 standard deviation units higher than control programs. When we control for covariates, the effect increases to 1.35 standard deviation units ($g = 1.05$). This suggests PWS-PMI has a moderate to large effect on classroom-level implementation fidelity of Pyramid Model practices. On the covariate-adjusted results for each of the CLASS measures (columns 4, 6, and 8), we find that PWS-PMI has a positive and significant effect of 1.55 standard deviations ($g = .26$) on instructional support, a positive and significant effect of 1.90 standard deviations ($g = .30$) on emotional support, and a positive and significant effect of 1.77 standard deviations ($g = .40$) on classroom organization.

For child-level outcomes (social skills and problem behavior), we again do not show the null models for brevity, but we use ICCs from null models to assess variability of these two outcomes at the classroom and program level. The classroom-level ICC shows that about 37.5% of the variance in social skills is at the classroom level, about 7.8% of the variance is at the program level, and the remainder is at the child level. For problem behaviors, the classroom and program level ICCs are 26.2% and 10.2%, respectively. Although the program level ICCs are not as high as those at the classroom level, conventional interpretations of ICC suggest that these values (all above 0.05) suggest medium to large group effects, supporting our decision to use three-level models (Bliese, 1998; LeBreton & Senter, 2008).

Table 5 shows the child-level results from our HLM models. Columns 1-2 show results where the outcome is the social skills measure from SSIS, and columns 3-4 show results for the SSIS measure of problem behavior. Controlling for the outcome at baseline and all covariates, the coefficient on social skills is positive (0.63 standard deviation units) but not statistically significant. The coefficient on problem behavior is 0.41 standard deviation units but also not statistically significant.

We tested these effect estimates for robustness to inclusion of alternative sets of covariates. When attempting to run the analysis including all available covariates, several covariates are perfectly collinear and multiple regressors are dropped due to limited degrees of freedom available for covariates, which were measured at the program level. Therefore, we sequentially test models that include all the covariates for which estimates are obtainable, and separately for program-level covariates, classroom-level, and child-level covariates. The results with alternative covariates are shown in Appendix Tables A6 and A7 and are consistent with the ef-

Table 3
Effect of PWS PMI Intervention on Program Level Measures: PreSET, SPIFI, and Disciplinary Actions.

	(1) SPIFISum of All Indicators	(2) SPIFISum of All Indicators	(3) SPIFISum of All Items	(4) SPIFISum of All Items	(5) PreSET	(6) PreSET	(7) Total Number of Student-Disciplinary Actions	(8) Total Number of Student-Disciplinary Actions
PWS-PMI	1.62***	1.63**	1.44**	1.44**	1.25*	1.28**	-0.14	-0.08
	(0.24)	(0.25)	(0.34)	(0.29)	(0.35)	(0.29)	(1.06)	(0.71)
	[0.00]	[0.00]	[0.01]	[0.00]	[0.01]	[0.01]	[0.90]	[0.92]
Baseline Measure		0.21		0.47		1.36		-0.52*
		(0.31)		(0.26)		(0.71)		(0.16)
		[0.53]		[0.14]		[0.11]		[0.03]
Observations	15	15	15	15	15	15	15	14
R^2 Overall	0.74	0.81	0.75	0.79	0.68	0.70	0.71	0.81
Block Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses. p-values in brackets. The small number of programs means that many covariates do not vary within blocks. Therefore, these models do not include covariates besides the baseline measure of the outcome.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table 4
Effect of PWS PMI Intervention on Classroom Level Measures: TPOT and CLASS.

	(1) TPOT	(2) TPOT	(3) Instructional Support	(4) Instructional Support	(5) Emotional Support	(6) Emotional Support	(7) Classroom Organization	(8) Classroom Organization
PWS-PMI	0.74***	1.35*	0.29	1.55*	-0.01	1.90**	0.10	1.77*
	(0.22)	(0.53)	(0.29)	(0.65)	(0.31)	(0.65)	(0.27)	(0.75)
	[0.00]	[0.01]	[0.32]	[0.02]	[0.98]	[0.00]	[0.72]	[0.02]
Outcome at Baseline	0.21	0.11	-0.06	-0.01	0.52*	0.42*	0.15	0.06
	(0.15)	(0.13)	(0.18)	(0.16)	(0.22)	(0.19)	(0.24)	(0.22)
	[0.15]	[0.39]	[0.72]	[0.96]	[0.02]	[0.03]	[0.54]	[0.80]
Tuition Fee		-0.55		-0.87		-1.22		-0.70
		(0.88)		(1.08)		(1.08)		(1.26)
		[0.53]		[0.42]		[0.26]		[0.58]
Head Start/EHS		-0.89		-2.20*		-2.20*		-3.18*
		(0.90)		(1.10)		(1.10)		(1.26)
		[0.32]		[0.05]		[0.05]		[0.01]
External Behavior Support		2.17		3.60*		4.36**		4.64**
		(1.17)		(1.40)		(1.43)		(1.63)
		[0.06]		[0.01]		[0.00]		[0.00]
Proportion Black Students		-0.48		5.67		5.45		6.23
		(3.32)		(3.99)		(4.08)		(4.62)
		[0.89]		[0.15]		[0.18]		[0.18]
Proportion White Students		-0.10		0.41		3.06		1.23
		(1.94)		(2.33)		(2.42)		(2.80)
		[0.96]		[0.86]		[0.20]		[0.66]
Proportion Black Teachers		1.92		-1.65		0.47		0.67
		(0.99)		(1.15)		(1.20)		(1.36)
		[0.05]		[0.15]		[0.69]		[0.62]
Proportion Teachers with BA or Above		0.49		-1.00		-2.21*		-1.52
		(0.90)		(1.09)		(1.12)		(1.31)
		[0.59]		[0.36]		[0.05]		[0.25]
Observations	40	40	40	40	40	40	40	40
Classroom ICC	4.13%	<1%	11.71%	<1%	16.02%	<1%	<1%	<1%
Block Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

EHS: Early Head Start; IFSP: Individualized Family Support Plan.
Standard errors in parentheses. p-values in brackets.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

fect estimates from our preferred specification, which we report above.

We attempted to further examine the extent to which missing data due to case-wise deletion altered the effect estimates. The most straightforward way to conduct this examination is to impute the missing data and reanalyze the data. However, the number of covariates available for the imputation was limited and the size of the sample relatively small for implementing multiple imputation. Appendix Tables A4 and A5 show the results from analyses on the full data set after implementing multiple imputation in two ways. Table A4 imputes missing data using the full sample, while Table A5 imputes variable values separately for the treatment and control groups. The coefficients estimated from the imputed data continue to be positive and are similar in magnitude to our pri-

Table 5
Effect of PWS PMI Intervention on Child Level SSIS Measures.

	(1) Social Skills	(2) Social Skills	(3) Problem Behaviors	(4) Problem Behaviors
PWS-PMI	0.11	0.63	-0.20	0.41
	(0.16)	(0.34)**	(0.13)	(0.28)
	[0.47]	[0.07]	[0.12]	[0.14]
Outcome at Baseline	0.52***	0.53***	0.55***	0.59***
	(0.05)	(0.05)	(0.06)	(0.06)
	[0.00]	[0.00]	[0.00]	[0.00]
Tuition Fee		-0.54		-0.48
		(0.57)		(0.50)
		[0.35]		[0.33]
Head Start/EHS		-0.92		-1.41*
		(0.67)		(0.56)
		[0.17]		[0.01]
External Support with Behavior		0.92		1.74*
		(0.83)		(0.68)
		[0.27]		[0.01]
Proportion Black Students		1.30		2.10
		(2.31)		(1.88)
		[0.57]		[0.27]
Proportion White Students		-0.57		-0.34
		(1.34)		(1.15)
		[0.67]		[0.77]
Proportion Black Teachers		-0.92		-0.18
		(0.71)		(0.56)
		[0.20]		[0.75]
Proportion Teachers with BA or Above		-0.76		-0.74
		(0.61)		(0.49)
		[0.21]		[0.13]
Observations	335	335	335	335
Classroom ICC	19.66%	11.35%	8.283%	1.003%
Program ICC	<1%	<1%	<1%	<1%
Block Fixed Effect	Yes	Yes	Yes	Yes

EHS: Early Head Start; IFSP: Individualized Family Support Plan.
Standard errors in parentheses. p-values in brackets.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

mary results, but for the classroom-level outcomes, the estimates are no longer statistically significant. The multiple imputation analyses make us more cautious about the original effect estimates, but the plausibility of the original estimates remains due to limitations on the covariates available for use in the imputation and the sample size.

6. Discussion

Our findings were multi-faceted. First, we found the SPIFI to be a useful and reliable tool for measuring the fidelity of treatment for PSW-PMI. Second, we found that the intervention, including training and coaching, was effective for supporting programs to use PWS-PMI. These effects were unequivocal and promising. Third, classroom effects were consistently positive, and child effects were in the right direction and viewed as promising. We discuss our outcomes and implications of each below.

Our study was the first randomized experiment to evaluate PWS-PMI. We found that training and support increased implementation of the program-wide practices needed for supporting the use of the Pyramid Model. The programs in the control group—who did not receive training and coaching—did not implement the Pyramid Model to the same extent. The differences across the two groups were meaningful in magnitude and statistically significant. However, programs did not reach what we consider to be high fidelity implementation as they did not reach full implementation of most of the indicators on the SPIFI. This has at least four implications for future research and practice.

First, the SPIFI provided a valid and reliable measure of PWS-PMI. The SPIFI was useful for measuring changes in program-wide implementation, which we refer to as intervention fidelity and measured critical components of program-wide implementation. Use of this program-level measure of treatment take-up may be useful in future studies of Pyramid Model interventions to better understand the treatment contrast between treatment and control sites and to reduce the plausibility of program-wide supports as an explanation if expected effects of a Pyramid intervention are not observed in practice. Also, it may be useful in practice as a tool to monitor program supports for Pyramid Model implementation. However, additional research is needed to identify what scores on the SPIFI are related to improved and meaningful classroom and child outcomes. We believe additional replications should be conducted to identify both the criterion score (i.e., percentage score on the SPIFI) and critical components (i.e., SPIFI indicators) related to meaningful classroom improvements and child social-emotional competence. More research is needed examining the relations between program-wide implementation, classroom quality, and child social competence.

Second, we provided training and coaching over a 10 month period with each program in the intervention group. Despite seeing strong improvements in PWS-PMI, implementation science research suggests that full implementation of PWS-PMI would require multiple years of training and support (Durlak, 2015; Fox et al., 2019). In this study, classroom coaches provided substantially lower doses of coaching than we anticipated and lower than would be needed to see classroom improvements (Hemmeter et al., 2016; Hemmeter et al., 2018). The gap between intended coaching and the intensity of coaching needed to establish practice implementation fidelity has been noted by researchers as a critical challenge (Downer et al., 2009; Weber-Mayrer et al., 2018). Fur-

ther, classroom coaching might be a critical component of PWS-PMI. There are several reasons why the dose was lower, including lack of confidence among coaches and inadequate time for coaching. To address this issue in future research, external coaches might focus on supporting teachers in the first year and then fade that support during the second year to build the capacity of the classroom coach to support sustained, high-fidelity implementation of Pyramid Model practices. It will be necessary to work with leadership teams to ensure coaches have adequate time to provide the dose of coaching needed to see change in teacher practice.

Third, though not statistically significant, the SSIS results suggested that children's challenging behavior scores may have increased over time in intervention programs. Given this is a teacher-completed measure, we hypothesize that this difference could be an artifact of teachers in the treatment group having a better understanding of the meaning of challenging behavior as a result of the intervention. That is, the PWS-PMI training and coaching increased teachers' understanding and awareness of the meaning of children's challenging behaviors. Additionally, our findings align with previous research showing increased SSIS problem behavior among older age groups throughout the school year (Schonert-Reichl & Lawlor, 2010). Additional research is needed examining teachers' perceptions of challenging behavior over at least two years of implementation of PWS-PMI. In addition, future research that examines the implementation of PWS-PMI over time should include a direct assessment or observational measure of children's social, emotional, and behavioral outcomes.

The child care system is multifaceted and complex. We included programs that served families experiencing many different types of challenges. Further, the child care programs had limited resources and external supports. We had higher than expected rates of child attrition and difficulties contacting families in both the intervention and control groups. Child care programs also reported having high levels of teacher and staff turnover, crises, and even closures. Some of these issues are systemic (e.g., staff turnover, limited resources) and point to a need for broader changes in our child care system. Future research should identify readiness markers or program characteristics that indicate a program is ready for PWS-PMI. These could be used to develop specific inclusion criteria for programs in research. The readiness markers also could be used to help programs identify their own readiness for PWS-PMI.

Finally, our goal was to build the capacity of child care programs to provide an implementation infrastructure for the implementation of evidence-based practices. We sought to establish the organizational and system components within the program that would contribute to the sustainable and high-fidelity implementation of practices time (Franks & Schroeder, 2013). This study provides promising data about the feasibility of a program-wide approach to implementation of an intervention and the potential outcomes for teachers and children.

6.1. Limitations

While this study is one of the first to examine program level implementation of social-emotional interventions in early childhood settings, there are limitations that should be considered in examining the results and which provide the foundation for future research. First, the study was limited in terms of the number of programs that were included, and this issue was further impacted by one program closing and ongoing attrition of staff at programs that remained in the study. While the effects of the attrition cannot be fully known, we conducted several tests for potential effects and did not find unusually high levels of attrition given the study population, differential attrition or compositional changes in the treatment and control samples over time. The results of multiple imputation raised a caution about the plausibility of the study findings. The lack of significant differences in the effect estimates for classroom-level outcomes in the imputed sample may be attributable to the lack of effect on the teachers and students who attrited, which could indicate bias or reflect low dosage and duration of exposure to PWS-PMI. Alternatively, it may indicate the covariates available for the multiple imputation were too limited to impute the missing data reliably. Ultimately, the effects of attrition can only be resolved through replication of the study.

Second, one of our goals was to examine the effects of the intervention on programs' inappropriate disciplinary actions. While the estimated coefficient on the reduction of disciplinary actions is sizeable, the lack of statistical significance may be attributable to the statistical power to detect effects at the program level. As noted above, we constructed several alternative measures of disciplinary outcomes at the program level, all of which led to similar conclusions. Recent research found that the Pyramid Model is associated with reductions in use of exclusionary discipline practices (Clayback & Hemmeter, 2021). However, in that study, the programs were situated in a state system that had been providing support for program-wide implementation of the Pyramid Model for several years. It is likely that these outcomes will take intensive support provided over a longer period of time than was provided in the current study.

7. Conclusion

PWS-PMI reflects a systematic approach to supporting early childhood programs in using Pyramid Model practices and enhancing social-emotional outcomes and is grounded in an understanding of implementation science. We found the SPIFI to be an effective, reliable, and useful tool to measure the implementation of PWS-PMI. Despite existing complexities and challenges, programs increased their implementation of PWS-PMI and showed strong promise for improving classroom practices and child outcomes after only one year of support. Our findings indicate a program-wide approach is effective and feasible in early childhood programs; however, more time and support may be necessary to sustain high fidelity implementation and more robust effects on children.

Authors' Statement

Mary Louise Hemmeter: Conceptualization, Methodology, Writing – Original Draft, Review and Editing, Funding Acquisition. Erin Barton: Conceptualization, Methodology, Writing – Original Draft, Review and Editing, Funding Acquisition. Lise Fox: Conceptualization, Methodology, Writing – Original Draft, Review and Editing, Funding Acquisition. Chris Vatland: Conceptualization, Methodology, Writing – Original Draft, Review and Editing, Funding Acquisition. Gary Henry: Methodology, Formal Analysis, Conceptualization, Writing – Original Draft, Review and Editing. Lam Pham: Methodology, Formal Analysis, Conceptualization, Writing – Original Draft, Review and Editing. Kymberly Horth: Investigation, Project Administration, Writing – Review & Editing. Abby Taylor: Investigation, Project Administration, Writing – Review & Editing. Denise Binder, Investigation, Project Administration, Writing – Review & Editing. Meghan von der Embse, Investigation, Project Administration, Writing – Review & Editing. Myrna Veguilla: Investigation, Project Administration, Writing – Review & Editing

Acknowledgments

The authors wish to acknowledge the programs, teachers, and children who participated in the study, the administrators who supported the study, and our staff and students who supported this work. This work was supported by a grant from the National Center for Education Research, Institute of Education Sciences, U.S.

Department of Education to Vanderbilt University (R305A150141). The opinions expressed are those of the authors, not the funding agency.

Disclosures

Drs. Hemmeter and Fox are authors of the Teaching Pyramid Observation Tool and receive a portion of royalties.

Appendix

Table A1,A2,A3,A4,A5,A6,A7

Table A1
Training Evaluation and Fidelity Data.

Training	Fall - Intervention Trainings % Fidelity(range)	*M* Evaluation Score[1] (range)	Spring – Control Trainings % Fidelity[2]	*M* Evaluation Score[1] (range)
LT	97 (96-98)	3.8 (3-4)	98	3.7 (1-4)
PBC	93 (93-93)	3.8 (2.5-4)	98	3.7 (2-4)
Staff	95 (75-100)	3.7 (1-4)	100	3.6 (1-4)
PTRYC	100 –	3.5 (2.5-4)	91	3.1 (1-4)
TPOT	97 (94-100)	3.6 (2-4)	100	3.4 (1-4)

LT: Leadership Team; PBC: Practice-Based Coaching; PTRYC: Prevent Teach Reinforce for Young Children; TPOT: Teaching Pyramid Observation Tool.
[1] Evaluation scale for questions 1-4
[2] Trainings occurred only at one site, thus there are no ranges for these fidelity scores.

Table A2
External Coaching Supports During the Study.

		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Average (range)
Leadership Team Meetings		9	10	10	10	9	9	9	6	9.0 (6-10)
Leadership Team Meeting Activities	Work on ECBoQ	3	4	4	4	3	6	1	3	3.5 (1-6)
	Work on Implementation Plan	7	7	9	7	6	7	0	3	5.8 (0-9)
	BIR data support	6	0	7	5	3	3	0	1	3.1 (0-7)
	TPOT/ECBoQ/Buy-in data support	5	2	5	8	0	2	0	2	3.0 (0-8)
Weekly Visits		29	18	36	22	24	34	23	14	25.0 (14-36)
Weekly Visit Activities	Discuss coaching	10	10	11	12	10	15	5	7	10.0 (5-12)
	Coach in classroom with coach	4	2	6	8	5	9	0	2	4.5 (0-9)
	Coach in classroom without coach	1	5	5	1	5	15	14	4	6.3 (1-15)
	Conduct TPOT with coach	2	1	0	1	3	0	0	1	1 (0-2)
	Behavior support in classroom	10	1	11	1	3	1	0	0	3.8 (0-11)
	Behavior support out of classroom	8	1	7	5	4	5	0	2	4.0 (0-8)
	Data support	12	3	13	10	6	5	1	4	6.8 (1-13)
	Family event	1	1	3	1	1	0	0	0	0.9 (0-3)

Coach in classroom with coach and Coach in classroom without coach are mutually exclusive. Work on ECBoQ = reviewing or updating the document; Work on Implementation Plan = reviewing or updating the document; BIR data support = reviewing the purpose, reviewing completed BIRs, assisting with entry, or interpreting data; Discuss coaching = talking with coach about coaching sessions or coaching in general, reviewing action plans and notes, giving suggestions, or reviewing video with the coach; Behavior support in classroom = observing or supporting implementation of a behavior support plan; Behavior support out of classroom = any other behavior support activities that occurred out of the classroom; Data support = suggesting data to collect, assisting with entering data, interpreting data, or creating Implementation Plan goals based on data for BIRs, TPOTs, ECBoQ, or buy-in data

Table A3
Descriptive Averages at Baseline and Post-Intervention.

	Baseline		Post-Intervention			
	M	*SD*	*M*	*SD*	Difference	*p*-value
Program Characteristics						
Tuition Fee	0.69	(0.48)	0.80	(0.41)	0.11	0.49
Head Start/EHS	0.31	(0.48)	0.33	(0.49)	0.02	0.91
External Support with Behavior	0.50	(0.52)	0.80	(0.41)	0.30	0.09
Child Characteristics						
Average Age (Years)	3.91	(0.20)	4.33	(0.28)	0.41***	0.00
Proportion Girls	0.47	(0.09)	0.47	(0.15)	0.00	0.99
Proportion Black	0.42	(0.33)	0.43	(0.37)	0.01	0.93
Proportion White	0.30	(0.32)	0.34	(0.34)	0.04	0.76
Proportion IEP/IFSP	0.02	(0.02)	0.02	(0.03)	0.00	0.66
Proportion Dual Language Learner	0.15	(0.21)	0.12	(0.23)	-0.03	0.73
Proportion Child Care Subsidy	0.50	(0.36)	0.49	(0.35)	-0.01	0.93
Teacher Characteristics						
Proportion Female Teachers	1.00	(0.00)	1.00	(0.00)	0.00	.
Proportion Black Teachers	0.58	(0.44)	0.63	(0.42)	0.05	0.78
Proportion White Teachers	0.38	(0.44)	0.35	(0.45)	-0.03	0.86
Proportion Teachers with BA or Above	0.28	(0.42)	0.38	(0.42)	0.10	0.54
Proportion Teachers with ECE Degree	0.70	(0.35)	0.63	(0.38)	-0.07	0.61
Proportion Teachers with Pre-K License	0.43	(0.31)	0.44	(0.39)	0.01	0.97
Average Teacher Experience (Months)	143.06	(85.55)	149.60	(84.92)	6.54	0.84
Average Months of Experience in Position	45.12	(39.97)	54.29	(42.23)	9.17	0.56

EHS: Early Head Start; IEP: Individual Education Plan; IFSP: Individualized Family Support Plan; ECE: Early Childhood Education.
Standard deviation in parentheses. The social skills scale and problem behaviors scale have been standardized to have a mean zero and unit variance.
* p < 0.05 ** p < 0.01 *** p < 0.001.

Table A4
Effect of PSW PMI Treatment on Child level SSIS Measures and Classroom Level Measures: TPOT and CLASS After Multiple Imputation.

	(1) Social Skills	(2) Problem Behaviors	(3) TPOT	(4) Instructional Support	(5) Emotional Support	(6) Classroom Organization
PWS-PMI	0.13	0.18	0.65	0.62	0.91	0.67
	(0.23)	(0.28)	(0.59)	(0.71)	(0.75)	(0.78)
	[0.57]	[0.52]	[0.28]	[0.39]	[0.24]	[0.39]
Outcome at Baseline	0.58***	0.52***	0.13	0.01	0.27	-0.15
	(0.05)	(0.05)	(0.28)	(0.25)	(0.27)	(0.44)
	[0.00]	[0.00]	[0.65]	[0.95]	[0.32]	[0.74]
Tuition Fee	-0.33	-0.36	0.21	-0.14	-0.83	0.47
	(0.47)	(0.50)	(1.26)	(1.35)	(1.47)	(1.63)
	[0.48]	[0.47]	[0.87]	[0.92]	[0.57]	[0.78]
Head Start/EHS	-0.50	-0.85	-0.75	-1.53	-1.66	-2.56+
	(0.51)	(0.56)	(1.12)	(1.22)	(1.28)	(1.31)
	[0.33]	[0.13]	[0.50]	[0.21]	[0.20]	[0.05]
External Support with Behavior	0.31	0.70	1.28	2.68	3.89*	3.63*
	(0.66)	(0.68)	(1.59)	(1.66)	(1.91)	(1.81)
	[0.63]	[0.30]	[0.42]	[0.11]	[0.04]	[0.05]
Proportion Black Students	-0.98	0.59	-1.52	0.72	3.81	2.13
	(1.17)	(1.26)	(3.13)	(3.40)	(3.94)	(3.76)
	[0.41]	[0.64]	[0.63]	[0.83]	[0.34]	[0.58]
Proportion White Students	-2.58**	-0.83	-2.03	-4.01	-0.19	-3.09
	(0.89)	(1.11)	(2.29)	(2.62)	(3.03)	(3.30)
	[0.00]	[0.46]	[0.38]	[0.14]	[0.95]	[0.36]
Proportion Black Teachers	-0.78	-0.14	0.95	-1.45	-0.57	0.43
	(0.52)	(0.59)	(1.32)	(1.47)	(1.55)	(1.38)
	[0.14]	[0.81]	[0.48]	[0.33]	[0.71]	[0.76]
Proportion Teachers with BA or Above	0.36	-0.05	1.47	0.91	-0.54	0.61
	(0.50)	(0.59)	(1.10)	(1.20)	(1.52)	(1.62)
	[0.48]	[0.93]	[0.19]	[0.45]	[0.72]	[0.71]
Observations	603 children	603 children	60 classrooms	60 classrooms	60 classrooms	60 classrooms
Block Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses. p-values in brackets. These analyses derive from all observations of students and classrooms after multiple imputation of missing values. The imputation process utilized multivariate normal regressions with a burn-in period of 100, producing 10 imputed datasets that were all used in the regression models. These observations include students and teachers who were observed (1) only at baseline, (2) only post-intervention, and (3) both at baseline and post-intervention. Multiple imputation could not be conducted for the program level analyses, because the small number of programs (16) was too few to reliably specify imputation models.
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

Table A5
Effect of PSW PMI Treatment on Child level SSIS Measures and Classroom Level Measures: TPOT and CLASS After Multiple Imputation (Treatment and Control Groups Imputed Separately).

	(1) Social Skills	(2) Problem Behaviors	(3) TPOT	(4) Instructional Support	(5) Emotional Support	(6) Classroom Organization
PWS-PMI	0.34*	0.15	0.51	0.76	0.06	0.12
	(0.13)	(0.19)	(0.28)	(0.50)	(0.82)	(0.96)
	[0.01]	[0.43]	[0.07]	[0.13]	[0.95]	[0.90]
Outcome at Baseline	0.48***	0.38***	0.22	0.08	0.05	-0.63
	(0.05)	(0.06)	(0.21)	(0.41)	(0.74)	(1.03)
	[0.00]	[0.00]	[0.30]	[0.85]	[0.94]	[0.55]
Tuition Fee	0.28	0.46	0.53	-0.22	-0.82	-0.82
	(0.22)	(0.33)	(0.55)	(0.77)	(1.13)	(1.28)
	[0.21]	[0.17]	[0.34]	[0.78]	[0.47]	[0.52]
Head Start/EHS	0.38	0.30	0.43	-0.58	-0.51	-1.28
	(0.24)	(0.33)	(0.61)	(0.75)	(1.19)	(1.46)
	[0.11]	[0.36]	[0.49]	[0.44]	[0.67]	[0.39]
External Support with Behavior	-0.24	-0.60	0.24	0.50	1.63	1.69
	(0.23)	(0.37)	(0.56)	(0.77)	(1.70)	(1.85)
	[0.31]	[0.12]	[0.67]	[0.52]	[0.35]	[0.37]
Proportion Black Students	-1.27*	-1.06	0.07	0.31	2.62	2.26
	(0.59)	(1.03)	(1.49)	(1.43)	(2.48)	(3.05)
	[0.03]	[0.31]	[0.96]	[0.83]	[0.29]	[0.47]
Proportion White Students	-1.06**	-0.20	0.07	0.11	0.38	-0.21
	(0.36)	(0.57)	(0.95)	(1.29)	(1.92)	(2.10)
	[0.00]	[0.72]	[0.94]	[0.93]	[0.84]	[0.92]
Proportion Black Teachers	-1.00***	-0.68	-0.01	-1.17	-0.59	0.22
	(0.23)	(0.50)	(0.58)	(0.98)	(1.64)	(1.79)
	[0.00]	[0.20]	[0.98]	[0.24]	[0.72]	[0.90]
Proportion Teachers with BA or Above	-0.67**	-0.69	1.20	-0.27	0.03	-0.04
	(0.25)	(0.46)	(0.78)	(1.16)	(1.87)	(2.02)
	[0.01]	[0.14]	[0.13]	[0.81]	[0.99]	[0.98]
Observations	603	603	60	60	60	60
Block Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

EHS: Early Head Start; IFSP: Individualized Family Support Plan.
Standard errors in parentheses. p-values in brackets. These analyses derive from all observations of students and classrooms after multiple imputation of missing values. The imputation process utilized multivariate normal regressions with a burn-in period of 100, producing 10 imputed datasets that were all used in the regression models.
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$

Table A6
Effect of PWS PMI Intervention on Classroom Level Measures: TPOT and CLASS – With Additional Covariates.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	TPOT	TPOT	TPOT	TPOT	Instructional Support	Instructional Support	Instructional Support	Instructional Support	Emotional Support	Emotional Support	Emotional Support	Emotional Support	Classroom Organization	Classroom Organization	Classroom Organization	Classroom Organization
PWS-PMI	1.12***	1.05***	0.82**	1.45**	0.98**	0.67*	1.06**	1.40*	0.51	0.64*	0.70	2.15***	0.45*	0.38*	0.62**	0.86**
	(0.26)	(0.23)	(0.31)	(0.48)	(0.32)	(0.28)	(0.38)	(0.59)	(0.32)	(0.28)	(0.38)	(0.59)	(0.18)	(0.17)	(0.01)	(0.22)
	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.02]	[0.00]	[0.02]	[0.10]	[0.02]	[0.07]	[0.00]	[0.03]	[0.04]	[0.02]	[0.03]
Outcome at Baseline	0.18	0.11	0.11	0.11	-0.04	-0.01	-0.01	-0.01	0.46*	0.42*	0.42*	0.42*	0.18	0.06	0.06	0.06
	(0.13)	(0.13)	(0.13)	(0.13)	(0.17)	(0.16)	(0.16)	(0.16)	(0.20)	(0.19)	(0.19)	(0.19)	(0.24)	(0.22)	(0.22)	(0.22)
	[0.18]	[0.39]	[0.39]	[0.39]	[0.82]	[0.96]	[0.96]	[0.96]	[0.02]	[0.03]	[0.03]	[0.03]	[0.45]	[0.80]	[0.80]	[0.80]
Program Characteristics																
Tuition Fee	0.23			-3.03	0.49			0.62	-0.28			-7.40**	1.11			-2.17
	(0.49)			(2.13)	(0.60)			(2.56)	(0.60)			(2.64)	(0.88)			(3.13)
	[0.64]			[0.16]	[0.41]			[0.81]	[0.64]			[0.00]	[0.21]			[0.49]
Head Start/EHS	-0.02			-1.14	-0.75			0.00	-0.01			-5.27***	-0.13			-1.79
	(0.47)			(1.14)	(0.57)			(1.39)	(0.57)			(1.46)	(0.67)			(1.74)
	[0.96]			[0.32]	[0.19]			[1.00]	[0.98]			[0.00]	[0.85]			[0.30]
External Support with Behavior	1.32*			1.33	1.18			0.54	2.15***			-1.76	0.92			0.07
	(0.51)			(0.85)	(0.61)			(1.02)	(0.62)			(1.10)	(0.73)			(1.36)
	[0.01]			[0.12]	[0.06]			[0.60]	[0.00]			[0.11]	[0.21]			[0.96]
Student Characteristics																
Average Age (Years)		0.43		1.36		-2.23**		-0.99		-0.73		-1.29		-0.78		-0.41
		(0.63)		(0.84)		(0.76)		(0.91)		(0.77)		(0.99)		(0.88)		(1.18)
		[0.49]		[0.10]		[0.00]		[0.28]		[0.34]		[0.19]		[0.37]		[0.73]
Proportion Girls		7.88		9.92*		-3.25		3.01		9.55		20.54***		7.41		12.20
		(4.09)		(4.55)		(4.88)		(5.34)		(4.98)		(5.46)		(5.66)		(6.26)
		[0.05]		[0.03]		[0.51]		[0.57]		[0.06]		[0.00]		[0.19]		[0.05]
Proportion White Students		2.44**		1.37		1.32		-1.15		1.67		7.28**		0.78		1.42
		(0.79)		(1.89)		(0.96)		(2.32)		(0.98)		(2.42)		(1.11)		(2.94)
		[0.00]		[0.47]		[0.17]		[0.62]		[0.09]		[0.00]		[0.49]		[0.63]
Proportion Individualized Education Plan/IFSP		33.36		50.96		-36.23		6.16		59.40		195.81**		19.28		70.19
		(30.63)		(47.95)		(36.80)		(57.71)		(37.51)		(60.60)		(43.34)		(71.83)

(*continued on next page*)

Table A6 (*continued*)

	(1) TPOT	(2) TPOT	(3) TPOT	(4) TPOT	(5) Instructional Support	(6) Instructional Support	(7) Instructional Support	(8) Instructional Support	(9) Emotional Support	(10) Emotional Support	(11) Emotional Support	(12) Emotional Support	(13) Classroom Organization	(14) Classroom Organization	(15) Classroom Organization	(16) Classroom Organization
		[0.28]		[0.29]		[0.32]		[0.92]		[0.11]		[0.00]		[0.66]		[0.33]
Proportion Dual Language Learner		-0.63		-6.61		7.48*		2.04		-1.33		-12.21*		1.21		-4.67
		(3.10)		(4.38)		(3.72)		(5.05)		(3.84)		(5.15)		(4.44)		(5.88)
		[0.84]		[0.13]		[0.04]		[0.69]		[0.73]		[0.02]		[0.79]		[0.43]
Proportion Child Care Subsidy		-0.08		0.68		1.89		-1.08		-3.94*		3.21**		-1.73		-0.18
		(1.50)		(0.93)		(1.85)		(1.13)		(1.87)		(1.19)		(2.24)		(1.38)
		[0.96]		[0.46]		[0.31]		[0.34]		[0.03]		[0.01]		[0.44]		[0.89]
Teacher Characteristics																
Proportion Black Teachers			-15.12	-0.85			-26.32**	-0.51			-30.21**	-1.33			-33.54**	-0.68
			(7.84)	(1.06)			(9.42)	(1.20)			(9.61)	(1.24)			(10.99)	(1.42)
			[0.05]	[0.42]			[0.01]	[0.67]			[0.00]	[0.28]			[0.00]	[0.63]
Proportion White Teachers			-14.07*	-0.75			-19.95*	0.86			-26.14**	-2.74*			-26.72**	-0.77
			(6.50)	(0.89)			(7.78)	(1.06)			(7.96)	(1.07)			(9.07)	(1.23)
			[0.03]	[0.40]			[0.01]	[0.42]			[0.00]	[0.01]			[0.00]	[0.53]
Proportion Teachers with BA or Above			11.41*	0.00			12.86*	0.00			18.87**	0.00			17.36**	0.00
			(4.72)	(.)			(5.65)	(.)			(5.78)	(.)			(6.58)	(.)
			[0.02]	[.]			[0.02]	[.]			[0.00]	[.]			[0.01]	[.]
Proportion Teachers with Early Childhood Degree			-2.50	0.00			-2.86	0.00			-3.35	0.00			-5.20*	0.00
			(1.79)	(.)			(2.17)	(.)			(2.20)	(.)			(2.53)	(.)
			[0.16]	[.]			[0.19]	[.]			[0.13]	[.]			[0.04]	[.]
Average Months of Experience in Position			0.04	0.00			0.16*	0.00			0.12	0.00			0.20*	0.00
			(0.06)	(.)			(0.07)	(.)			(0.07)	(.)			(0.08)	(.)
			[0.44]	[.]			[0.02]	[.]			[0.09]	[.]			[0.01]	[.]
Observations	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
Classroom ICC	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%
Block Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

EHS: Early Head Start; IFSP: Individualized Family Support Plan.
Standard errors in parentheses. p-values in brackets.
* $p < 0.05$;
** $p < 0.01$;
*** $p < 0.001$.

Table A7
Effect of PSW PMI Treatment on Student Level SSIS Measures – With Additional Covariates

	(1) Social Skills	(2) Social Skills	(3) Social Skills	(4) Social Skills	(5) Problem Behaviors	(6) Problem Behaviors	(7) Problem Behaviors	(8) Problem Behaviors
PWS-PMI	0.48	0.27	0.46	0.54	0.17	0.18	0.14	0.03
	(0.27)	(0.17)	(0.23)	(0.39)	(0.16)	(0.14)	(0.18)	(0.14)
	[0.08]	[0.12]	[0.05]	[0.23]	[0.31]	[0.21]	[0.45]	[0.83]
Outcome at Baseline	0.50***	0.53***	0.53***	0.53***	0.56***	0.59***	0.59***	0.59***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Program Characteristics								
Tuition Fee	0.20			51.45	-0.67			42.02
	(0.64)			(29.84)	(0.54)			(24.64)
	[0.75]			[0.08]	[0.21]			[0.09]
Head Start/EHS	-0.53			-1.35*	-0.58			-0.99*
	(0.47)			(0.56)	(0.40)			(0.47)
	[0.26]			[0.02]	[0.15]			[0.04]
External Support with Behavior	0.29			1.36	0.06			1.36
	(0.49)			(0.86)	(0.41)			(0.71)
	[0.55]			[0.11]	[0.89]			[0.06]
Student Characteristics								
Average Age (Years)		-1.11*		-0.35		-0.82*		-0.13
		(0.44)		(0.47)		(0.32)		(0.37)
		[0.01]		[0.46]		[0.01]		[0.73]
Proportion Girls		-4.35		-2.06		-0.78		0.31
		(2.34)		(1.56)		(2.00)		(1.28)
		[0.06]		[0.19]		[0.70]		[0.81]
Proportion White Students		-0.58		0.00		-0.19		0.00
		(0.59)		(.)		(0.50)		(.)
		[0.33]		[.]		[0.71]		[.]
Proportion Individualized Education Plan/IFSP		-30.02		0.00		-16.55		0.00
		(17.88)		(.)		(15.80)		(.)
		[0.09]		[.]		[0.29]		[.]
Proportion Dual Language Learner		5.49*		0.00		4.24*		0.00
		(2.24)		(.)		(1.87)		(.)
		[0.01]		[.]		[0.02]		[.]
Teacher Characteristics								
Proportion Black Teachers			-8.28	0.00			-13.62**	0.00
			(5.61)	(.)			(4.53)	(.)
			[0.14]	[.]			[0.00]	[.]
Proportion White Teachers			-4.91	0.00			-9.79*	0.00
			(4.64)	(.)			(3.81)	(.)
			[0.29]	[.]			[0.01]	[.]
Proportion Teachers with BA or Above			1.68	0.00			5.31	0.00
			(3.40)	(.)			(2.87)	(.)
			[0.62]	[.]			[0.06]	[.]
Proportion Teachers with Early Childhood Degree			-1.63	0.00			-2.51*	0.00
			(1.43)	(.)			(1.22)	(.)
			[0.25]	[.]			[0.04]	[.]
Average Months of Experience in Position			0.07	0.00			0.10**	0.00
			(0.04)	(.)			(0.04)	(.)
			[0.12]	[.]			[0.01]	[.]
Observations	335	335	335	335	335	335	335	335
Classroom ICC	15.03%	11.35%	11.35%	11.35	3.46%	1.00%	1.00%	1.00%
Program ICC	<1%	<1%	<1%	<1%	<1%	<1%	<1%	<1%
Block Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

EHS: Early Head Start; IFSP: Individualized Family Support Plan.
Standard errors in parentheses. p-values in brackets. Proportion child care subsidy omitted because of collinearity.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

References

Barton, E. E., Steed, E. A., Strain, P., Dunlap, G., Powell, D., & Payne, C. J. (2014). An analysis of classroom-based and parent-focused social-emotional programs for young children. *Infants & Young Children, 27*(1), 3–29.

Bierman, K. L., Motamedi, M., & Domitrovich, C. (2015). Social and emotional learning programs for preschool children. In J. Durlak, R. P. Weissberg, & T. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 135–151). Guilford.

Blair, K. C., Lee, I., Cho, S., & Dunlap, G. (2011). Positive behavior support through family–school collaboration for young Children with Autism. *Topics in Early Childhood Special Education, 31*(1), 22–36. https://doi.org/10.1177/0271121410377510.

Bliese, P. D. (1998). Group size, ICC values, and group-level correlations: A simulation. *Organizational Research Methods, 1*(4), 355–373. https://doi.org/10.1177/109442819814001.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25*(2), 166–176.

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

Clayback, K. A., & Hemmeter, M. L. (2021). Exclusionary discipline practices in early childhood settings: A survey of child care directors. *Early Childhood Research Quarterly, 55*, 129–136. https://doi.org/10.1016/j.ecresq.2020.11.002.

Conroy, M. A., Brown, W. H., & Olive, M. L (2008). Social competence interventions for young children with challenging behavior. In W. H. Brown, S. L. Odom, & S. R. McConnell (Eds.), *Social competence of young children. Risk, disability, and intervention* (pp. 205–232). Brookes.

Covington-Smith, S., Lewis, T. J., & Stormont, M. (2011). The effectiveness of two universal behavioral supports for children with externalizing behavior in Head Start classrooms. *Journal of Positive Behavior Interventions, 13*(3), 133–143.

Domitrovich, C. E., & Greenberg, M. T. (2000). The study of implementation: Current findings from effective programs that prevent mental disorders in school-age children. *Journal of Educational and Psychological Consultations, 11*(2), 193–221.

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.

Domitrovich, C. E., Moore, J. E., & Greenberg, M. T. (2012). Maximizing the effectiveness of social-emotional interventions for young children through high-quality implementation of 3 evidence-based interventions. In B. Kelly, & D. F. Perkins (Eds.), *Handbook of implementation science for psychology in education* (pp. 207–229). Cambridge University Press.

Downer, J. T., Locasale-Crouch, J., Hamre, B., & Pianta, R. (2009). Educator characteristics associated with responsiveness and exposure to consultation and online professional development resources. *Early Education & Development, 20*, 431–455.

Dunlap, G., Strain, P., Lee, J. K., Joseph, J., & Leech, N. (2018). Randomized controlled evaluation of Prevent-Teach-Reinforce for young children. *Topics in Early Childhood Special Education, 37*(4), 195–205.

Dunlap, G., Wilson, K., Strain, P., & Lee, J. (2013). *Prevent-teach-reinforce for young children: The early childhood model of individualized positive behavior support*. Brookes.

Dunst, C. J., & Trivette, C. M. (2009). Using research evidence to inform and evaluate early childhood intervention practices. *Topics in Early Childhood Special Education, 29*(1), 40–52.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.

Durlak, J. A. (2015). What everyone should know about implementation. In J. A. Durlak, C. E Domitrovich, R. P. Weissberg, & T. P. Gullota (Eds.), *Handbook of Social Emotional Learning: Research and Practice* (pp. 395–405). The Guilford Press.

Fox, L., Lentini, R., & Binder, D. P. (2013). Promoting the social-emotional competence of all children: Implementing the Pyramid Model program-wide. *Young Exceptional Children Monograph Series No, 15*, 1–13.

Fox, L., Smith, B. J., & Law, D. P. (2019). How do we scale-up multi-tiered systems of support?. In J. Carta, & R. M. Young (Eds.), *Multi-Tiered Systems of Support: Driving change in early education* (pp. 215–234). Paul H. Brookes.

Franks, R. P., & Schroder, J. (2013). The key components of successful implementation. In T. Halle, A. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 5–20). Brookes Publishing.

Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist, 58*, 466–474.

Greenberg, M. T., Domitrovich, C. E., Weissberg, R. P., & Durlak, J. A. (2017). Social and emotional learning as a public health approach to education. *The Future of Children, 27*(1), 13–32.

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales manual*. Pearson Assessments.

Halle, T., Metz, A., & Martinez-Beck, I. (2013). *Applying implementation science in early childhood programs and systems*. Paul H. Brookes Publishing.

Hemmeter, M. L., Fox, L., & Snyder, P. (2013). A tiered model for promoting social-emotional competence and addressing challenging behavior. In V. Buysse, & E. Peisner-Feinberg (Eds.), *Handbook of response to intervention in early childhood* (pp. 85–101). Brookes.

Hemmeter, M. L., Fox, L., & Snyder, P. (2014). Teaching Pyramid Observation Tool—Research Edition *[Manual]*. Brookes.

Hemmeter, M. L., Fox, L., Snyder, P., Algina, J., Hardy, J., Bishop, C., & Veguilla, M. (2021a). Corollary Child Outcomes from the Pyramid Model Professional Development Intervention Efficacy Trial. *Early Childhood Research Quarterly, 54*, 204–218.

Hemmeter, M. L., Snyder, P. A., Fox, L., & Algina, J. (2016). Evaluating the implementation of the Pyramid Model for promoting social-emotional competence in early childhood classrooms. *Topics in Early Childhood Special Education, 36*(3), 133–146. https://doi.org/10.1177/0271121416653386.

Hemmeter, M. L., Ostrosky, M. M., & Fox, L. (2021b). *Unpacking the Pyramid Model: A Practical Guide for Preschool Teachers*. Baltimore, MD: Brookes Publishing Co.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel Analysis: Techniques and Applications* (3rd ed.). Routledge.

Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88–110.

McIntosh, K., Mercer, S. H., Hume, A. E., Frank, J. L., Turri, M. G., & Mathews, S. (2013). Factors related to sustained implementation of schoolwide positive behavior support. *Exceptional Children, 79*, 293–311.

Meek, S. E., & Gilliam., W.S. (2016). *Expulsion and Suspension as Matters of Social Justice and Health Equity*. Discussion Paper, National Academy of Medicine, Washington, DC. Available at: https://nam.edu/wp-content/uploads/2016/10/Expulsion-and-Suspension-in-EarlyEducation-as-Matters-of-Social-Justice-and-HealthEquity.pdf.

Metz, A., & Bartley, L. (2012). Active implementation frameworks for program success. *Zero to Three Journal, 32*(4), 11–18.

National Research Council. (2001). *Eager to learn: Educating our preschoolers*. National Academies Press.

Oberle, E., Domitrovich, C. E., Meyers, D. C., & Weissberg, R. P. (2016). Establishing systemic social and emotional learning approaches in schools: A framework for schoolwide implementation. *Cambridge Journal of Education, 46*, 277–297. https://doi.org/10.1080/0305764X.2015.1125450.

Odom, S. L. (2009). The ties that bind: Evidence-based practice, implementation science, and outcomes for children. *Topics in Early Childhood Special Education, 29*(1), 53–61.

Osher, D., Kidron, Y., Brackett, M., Dymnicki, A., Jones, S., & Weissberg, R. P. (2016). Advancing the science and practice of social and emotional learning: Looking back and moving forward. *Review of Research in Education, 40*, 644–681.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System [CLASS] manual: Pre-K*. Brookes.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed). Sage Publications.

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly, 26*(1), 1–13.

Schonert-Reichl, K. A., & Lawlor, M. S. (2010). The effects of a mindfulness-based education program on pre-and early adolescents' well-being and social and emotional competence. *Mindfulness, 1*(3), 137–151.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A Primer*. Sage Publications.

Snell, M. E., Berlin, R. A., Voorhees, M. D., Stanton-Chapman, T. L., & Hadden, S. (2012). A survey of preschool staff concerning problem behavior and its prevention in Head Start classrooms. *Journal of Positive Behavior Interventions, 14*(2), 98–107.

Snijders, T., & Bosker, R. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (2nd ed). Sage Publications.

Snyder, P. A., Hemmeter, M. L., Fox, L., Crowe Bishop, C., & Miller, M. D. (2013). Developing and gathering psychometric evidence for a fidelity instrument: The Teaching Pyramid Observation Tool—Pilot version. *Journal of Early Intervention, 35*(2), 150–172.

Snyder, P. A., Hemmeter, M. L., & Fox, L. (2015). Supporting implementation of evidence-based practices through practice-based coaching. *Topics in Early Childhood Special Education, 35*(3), 133–143.

StataCorp. (2015). *Stata Statistical Software: Release 14*. StataCorp LP.

Steed, E. A., & Webb, M. L. (2013). The psychometric properties of the preschool-wide evaluation tool (PreSET). *Journal of Positive Behavior Interventions, 15*(4), 231–241.

Steed, E. A., Pomerleau, T. M., & Horner, R. (2012). *Preschool-wide evaluation tool implementation manual, research edition*. Brookes.

U.S. Department of Education Office for Civil Rights. (2014). *Civil rights data collection. Data snapshot: School discipline*. Washington, DC: Author Available at: https://www2.ed.gov/about/offices/list/ocr/docs/crdc-discipline-snapshot.pdf .

U.S. Department of Education, Office of Civil Rights. (October 2016). 2013-2014 Civil Rights Data Collection. A First Look. Available at: https://www2.ed.gov/abou1t/offices/list/ocr/docs/2013-14-first-look.pdf

U.S. Department ofHealth and Human Services and U.S. Department of Education. (2014). *Policy statement on expulsion and suspension policies in early childhood settings*. Washington, DC: Author Available at https://www2.ed.gov/policy/gen/guid/schooldiscipline/policy-statement-ece-expulsions-suspensions.pdf .

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

Webster-Stratton, C., & McCoy, K. P. (2015). Bringing The Incredible Years® programs to scale. In K. P. McCoy & A. Diana (Eds.), The science, and art, of program dissemination: Strategies, successes, and challenges. New Directions for Child and Adolescent Development,149, 81–95.

Webster-Stratton, C., Reid, J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: evaluation of the Incredible Years Teacher and Child Training Programs in high-risk schools. *The Journal of Child Psychology and Psychiatry, 49*(5), 471–488. https://doi.org/10.1111/j.1469-7610.2007.01861.x.

Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management, 33*(3), 778–808.

Wenz-Gross, M., & Upshur, C. (2012). Implementing a primary prevention social skills intervention in urban preschools: Factors associated with quality and fidelity. *Early Education & Development, 23*(4), 427–450.

What Works Clearinghouse. (2013). *Assessing attrition bias*. Washington, DC: US Department of Education, Institute of Education Sciences.

What Works Clearinghouse. (2014). *Assessing attrition bias—Addendum (Version 3.0)*. Washington, DC: US Department of Education, Institute of Education Sciences.