Journal of School Psychology 91 (2022) 65–80

Contents lists available at ScienceDirect

Journal of School Psychology

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

Indirect effects of coaching on pre-K students' engagement and literacy skill as a function of improved teacher–student interaction[☆]

Robert C. Pianta [*], Daniel Lipscomb, Erik Ruzek

University of Virginia, United States of America

ARTICLE INFO

Action Editor: Linda Reddy

Keywords:
School readiness
Teacher coaching
Teacher-student interaction
Professional development interventions

ABSTRACT

The present study is a secondary analysis of a randomized control trial that had established the causal impact of MyTeachingPartner (MTP) coaching for improving the quality of teacher-student interactions. This study reports auxiliary analyses of the extent to which MTP was associated indirectly with classroom engagement and literacy outcomes for pre-kindergarten (pre-K) students as a function of the association of dosage of MTP cycles with teacher-student interactions. The number of coaching cycles a teacher completed was examined as a predictor of teacher-student interactions, as were corresponding direct or indirect associations with students' early literacy outcomes. The significant indirect effects detected provided support for the premise that coaching can improve student outcomes as a function of improvements in teachers' behaviors with students. When exposed to more cycles of feedback regarding interactions, teachers demonstrated improvements in their emotionally supportive interactions with students, which in turn predicted greater increases in students' positive classroom engagement. No significant indirect associations were detected for literacy outcomes. This linkage of coaching inputs, teacher-student interaction, and students' increased engagement confirms the hypothesized effects and theory of change associated with MTP and of coaching more generally.

Improving features of teachers' classroom practices that promote student learning is one emphasis of many efforts to increase the modest benefits of early education programs (e.g., Camilli et al., 2010; Pianta et al., 2021). For example, of the levers included in most policies or investments in teacher professional development, teachers' classroom interactions with students appear to be among the more predictive of students' school readiness (Sabol et al., 2013). Research has demonstrated the potency of teachers' general instructional behaviors and practices (Farran et al., 2011) and delivery of instructional content (Clements & Sarama, 2008; Pianta et al., 2021). Acknowledging this focus on the importance of teachers' classroom practice, over the past decade teacher professional development has shifted toward coaching (Herren, 2009; Snyder et al., 2015) as an approach to improve teacher practice (e.g.,

☆ We thank the generous programs and teachers who participated in the study reflected in the analyses reported here, as well as Marcia Kraft-Sayre, Sarah Hadden, Allison Leach, Grace Funk, Bridget Hamre, Andrew Mashburn, and Jason Downer for their contributions. The research reported herein was supported by the National Institute of Child Health and Human Development and the Interagency Consortium on School Readiness (NIH 5 R01 HD046061), to the University of Virginia. The opinions expressed are those of the authors and do not represent views of the National Institutes of Health.

* Corresponding author.
E-mail address: rcp4p@virginia.edu (R.C. Pianta).

https://doi.org/10.1016/j.jsp.2021.12.003
Received 4 February 2021; Received in revised form 8 September 2021; Accepted 16 December 2021
Available online 17 January 2022
0022-4405/

Diamond & Powell, 2011; Domitrovich et al., 2009; Pianta et al., 2008b) and student learning (Clements & Sarama, 2008; Downer et al., 2011).

The present study is a secondary analysis of a randomized control trial that established the causal impact of MyTeachingPartner (MTP) coaching for improving the quality of teacher-student interactions (Pianta et al., 2008b). In so doing, the present study examined the extent to which teachers' improvement in their classroom interactions with students as a function of variation in dosage of MTP coaching cycles (Mashburn et al., 2010; Pianta et al., 2008b) was related to gains in students' learning in the areas of classroom engagement and early literacy, two key features of school readiness, and attempted to confirm findings from a parallel analysis of indirect effects in a separate randomized evaluation trial of MTP in pre-K educational programs. In this prior study, variation in exposure to cycles of MTP coaching and feedback that targeted teachers' instructional support behaviors were related to improvements in teachers' instructional support interactions, which in turn predicted increased performance in children's literacy (Pianta et al., 2021). Of note is that the present study is not designed technically as a study in causal mediation, but rather it examines correlations reflecting the indirect paths, and causal chain, presumed to underlie all coaching models—that coaching fosters teachers' improved classroom practice, which in turn supports increased student learning.

Increasing student learning as a function of targeting teachers' classroom practice (particularly instruction) is the primary purpose of most professional development. Meta-analytic findings suggest that training early childhood educators on classroom practices improves both their skills ($d = 0.45$, $SE = 0.10$) and the outcomes of children they serve ($d = 0.55$, $SE = 0.30$), particularly when focused on specific educator behaviors (Fukkink, 2007). Conversely, coaching is not a well-specified model for professional development; the Snyder et al. (2012) review concluded that although coaching was popular, the lack of clear delineation of coaching models, mechanisms, and impacts made it difficult to identify coaching features that could drive scale-up efforts or further research. In subsequent writings, Snyder et al. (2015) specified components of "practice-based coaching" including observation, reflection, and feedback aligned to a definition of effective teaching practice, aspects of coaching that are included in the MTP approach that is the focus of the present study. In support of the presumed importance of coaching as a form of professional development in early childhood education, it was described by the Advisory Committee on Head Start Research and Evaluation as an evidence-supported approach for improving program quality and student outcomes (e.g., Lloyd & Modlin, 2012) in Head Start.

Several experimental studies have confirmed that coaching can lead to improvements in teacher practice, whether targeting implementation of a specific curriculum (Bierman et al., 2008; Clements et al., 2013; Domitrovich et al., 2009; Hsieh et al., 2009; Raver et al., 2008; Wasik & Hindman, 2011), or more general features of teacher-child interaction (Pianta et al., 2008b). Coaching that directly targets teachers' behavior appears highly effective in producing improvements in teachers' instructional, management, and socio-emotional behaviors and supports for children (Diamond & Powell, 2011; Domitrovich et al., 2009; Hsieh et al., 2009; Landry et al., 2009; Raver et al., 2008; Snyder et al., 2012). Although experimental studies demonstrate significant coaching effects on teacher practice, effect sizes of coaching on student outcomes are always smaller for student achievement (Powell et al., 2010; Raver et al., 2008) and social skills (Williford & Shelton, 2008). In sum, coaching that targets teachers' classroom practices appears to consistently improve those practices; what is less clear are the effects of coaching on student outcomes as a function of improved practice (Barton et al., 2011; Pianta et al., 2021; Shannon et al., 2015; Snyder et al., 2012).

1.1. MTP coaching, teacher behavior, and student learning

The present study draws its data from a randomized controlled evaluation of MTP in a sample of publicly funded pre-K classrooms. MTP coaching takes place through repeated cycles of internet-mediated interaction (both synchronous and asynchronous) between a teacher and coach. Teachers regularly videotape their instruction and send this video file to their coach, who edits the video into three brief segments that highlight a specific domain of teacher-student interaction (e.g., instructional support), as specified by the Classroom Assessment Scoring System (CLASS) observation framework (Pianta et al., 2008b). Each video segment is paired with accompanying annotation and is intended to (a) draw teachers' attention to (seeing) instances of their own effective behavior, (b) identify links between teacher behavior and engagement of students, and (c) analyze alternatives to observed behaviors. The three video segments and coaches' written feedback (annotation and questions that foster teachers' observation and analysis, i.e., "prompts") are posted to a secure website where, for each coaching cycle, teachers view segments and read and respond to prompts. Once teachers respond to coaches' prompts, the MTP cycle completes when teachers and coaches meet (via telephone or Internet) to analyze teachers' responses to the video segments and coaches' prompts, and plan applications to the classroom. The logic model for MTP effects is based on evidence suggesting that an organized framework for understanding the impact of interactions, practice in identifying and describing effective (and ineffective) interactions, and opportunities to see and analyze one's own interactions combine to form cognitive schema that guide teachers' practice and reduce cognitive overload (Hamre et al., 2014; Pianta et al., 2021). In the present study, we focused on variation in the number of coaching cycles as an index of the dosage of these coaching inputs to which teachers were exposed. We expected that a higher number of coaching cycles would be directly related to greater improvements in teacher-student interactions and indirectly to student outcomes.

Aligned with general theories of practice-focused coaching (e.g., Bierman et al., 2008; Hemmeter et al., 2015; Landry et al., 2009), MTP coaching uses the CLASS as the lens and definition for effective and ineffective teacher-child interaction and thus as the target for improvement. CLASS has proven to be predictive of children's school readiness outcomes (e.g., literacy, language skills, self-regulation; see Hamre et al., 2014); therefore, in the MTP model, it serves as a means of defining, identifying, and targeting teacher behaviors for improvement and as a basis for coaching and feedback to teachers (Pianta et al., 2017). The present study retained the CLASS three-factor structure (i.e., emotional support, organizational support, and instructional support) when examining and describing teacher-student interaction and its effects. Although there are alternative structures that have been advanced through bifactor modeling (e.g., Hamre et al., 2014) and meta-analytic work (see Li et al., 2019), because MTP coaching explicitly used the three-domain structure in written feedback to teachers and in more general information to teachers about interactions, the use of the three-domain structure is best aligned to the intent and effects of the intervention being considered.

MTP has been evaluated in two separate controlled trials in which teachers were assigned randomly to MTP or a business-as-usual control group. In the first evaluation of MTP, which produced the data set analyzed in the present study, coaches were professional educators hired for the purposes of the study and were based at the university research lab. The second evaluation, as part of the National Center for Research on Early Childhood Education (NCRECE), implemented MTP in 10 different urban pre-k programs and used local coaches identified by program directors.

In both evaluations of MTP, relative to control teachers, pre-K teachers assigned at random to MTP coaching improved their interactions and their students demonstrated greater gains in learning outcomes (Downer et al., 2012; Pianta et al., 2008b). In these evaluations of MTP, significant effect sizes (*d*) for impact on teacher-student interactions have ranged from 0.40 to 0.80 across various dimensions of the CLASS (Pianta et al., 2014). Regarding child outcomes, the first implementation and evaluation of MTP impacts noted significant effect sizes (*d*) ranging from 0.12 for literacy (Downer et al., 2012) to 0.27 for receptive language (Mashburn et al., 2010). Significant effects (*d*) for teacher-reported social outcomes were 0.21 for task orientation and 0.28 for assertive social skills (Hamre et al., 2012). The NCRECE multi-site evaluation of MTP (Pianta et al., 2017) detected significant positive effect sizes (*d*) on direct assessments of children's self-regulation/inhibitory control (0.24), teacher-reported persistence (0.30), and observed language productivity (0.21). Under conditions in which pre-K classrooms enrolled a higher percentage of 4-year-olds, MTP coaching was associated with significant gains in direct assessments of children's literacy skills and receptive language (Ansari & Pianta, 2018). Overall, these effect sizes for teacher practice and student outcomes are consistent with those reported from evaluations of other school-based professional development interventions that focus on teacher-student interaction (e.g., Bierman et al., 2008).

These prior comparisons of teachers in the present sample who were randomly assigned to MTP coaching and the control group have established the causal impact of MTP on (a) teacher-student interactions and (b) students' literacy, self-regulation, and social skills (Downer et al., 2011; Hamre et al., 2012; Mashburn et al., 2010; Pianta et al., 2008b). When indirect effects of MTP were examined in an auxiliary analysis of the NCRECE evaluation data set, the number of coaching cycles in which teachers received targeted feedback on their instructional interactions improved their instructional behavior, which in turn related to larger gains in students' literacy skills (Pianta et al., 2021). The number of coaching cycles was of interest in that study because MTP treatment group teachers differed in the number of cycles in which they engaged and there was a more general interest in treatment dosage and potential thresholds for impact that was informed by the literature on school-based interventions; moreover, dosage of an intervention tends to be the primary driver of cost (Sheridan et al., 2009). This prior finding suggests that MTP dosage (in that case the number of cycles) may warrant further investigation as an important parameter in planning implementation and in understanding treatment impact. The current study expands this focus on dosage and on indirect associations that reflect the causal logic of coaching in an auxiliary analysis of the dataset from the first (not NCRECE) implementation of MTP.

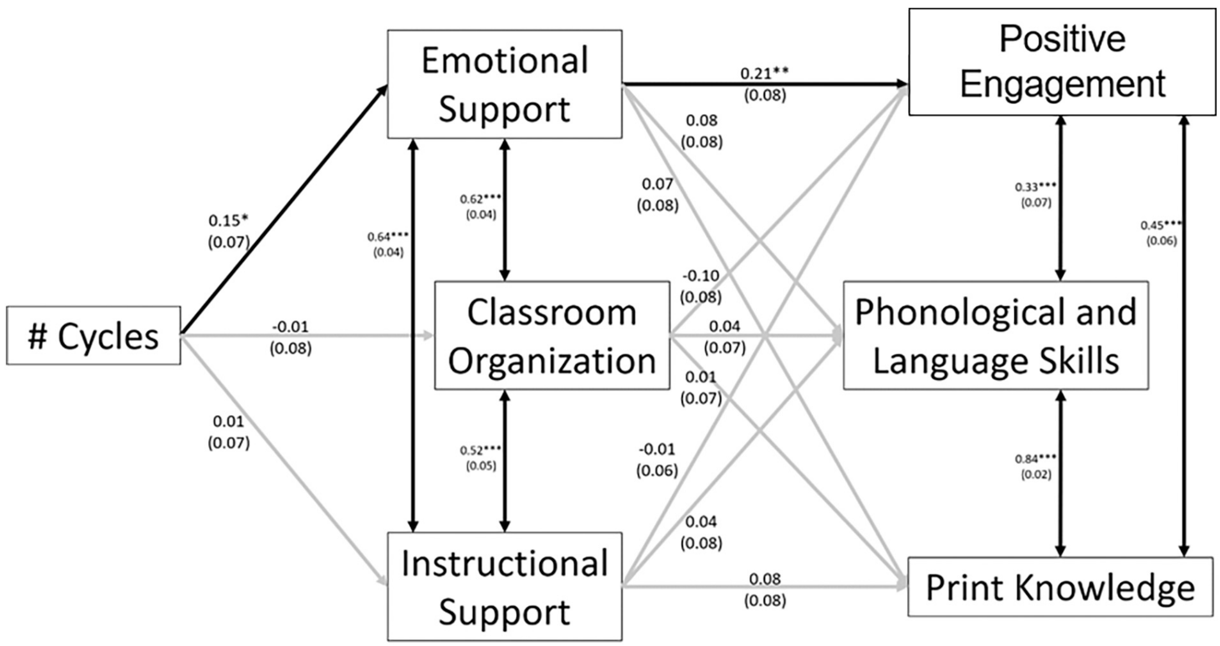

Fig. 1. SEM used to test the potential mediated effect of the number of intervention cycles on student factor outcomes, through teacher-student interaction quality (i.e., CLASS domains). *Note.* CFI = 0.99, TLI = 0.97, RMSEA = 0.04, SRMR = 0.03. Lines with single arrows represent regressions and double arrows showing covariances. Significant paths are in black and non-significant ones are in grey. Standardized coefficients *b* are presented adjacent to path lines. Covariates for CLASS domains and factor outcomes are presented in Table 5.

Finally, given that the feature of MTP to be examined in this study (e.g., number of coaching cycles) are present in other coaching models (Hemmeter et al., 2015), the results may have implications extending beyond the MTP model. For example, the conceptual framework outlined by Snyder et al. (2015) describes the ongoing and cyclical nature of coaching as a key feature in which coaches observe to gather and record information on practice and provide feedback on practice. In the current study, it was expected that the greater number of doses of coaching to which a teacher is exposed would relate to greater improvements in their practice. Findings related to this hypothesis would have implications for planning and understanding effects of coaching implementations.

1.2. The present study

Although the effects of MTP coaching described above have also been detected in less well controlled studies of its implementation (e.g., Early et al., 2017), unexplored in MTP evaluations, and in the coaching literature more broadly, are questions related to the indirect pathways through which coaching is responsible for improvements both in teachers' practices and student outcomes as well as the extent to which coaching dosage may be associated with such results. Sheridan et al. (2009) suggested that studies of professional development in early education be designed to better understand the "processes underlying effects"; in other words, studies evaluating professional development models should also attempt to identify processes that may be responsible for any effects (see Desimone, 2009). In the present study we investigated associations between the number of MTP coaching cycles and teacher-student interaction, and corresponding direct or indirect pathways associated with student outcomes in early literacy. For this analysis of MTP coaching as it may be related to gains in teacher and student outcomes, we evaluated the path model depicted in Fig. 1.

As noted in Fig. 1, the MTP component of interest is the number of cycles of feedback in which teachers were engaged. As was the case in the NCRECE examination of MTP indirect effects (Pianta et al., 2021), the present study included in the analysis all teachers in the sample, both treatment and control groups, where non-intervention teachers have zero MTP cycles and intervention teachers would have more than zero MTP cycles. This approach to secondary analysis is designed as an exploratory investigation of how variation in an element of a specific treatment (e.g., cycles of coaching) may relate to teacher and student outcomes; by including the control group, this exploratory analysis had increased power to detect associations that may be relevant for further investigation. The direct effects of MTP feedback cycles on student outcomes are evaluated as well as the indirect effects, as a function of features of teachers' interactions. For purposes of detecting indirect paths, analyses employed the standard (and moderately intercorrelated) CLASS domains (i.e., instructional, emotional, and organizational support). These domains of teacher-student interactions have been validated in factor-analytic work using multiple, large samples and shown to predict different areas of student learning (Hamre et al., 2012). Outcomes include key features of school readiness that are often the focus of public pre-kindergarten programs, including students' early literacy skills and teachers' ratings of students' positive engagement with teachers, peers, and tasks in the classroom. We hypothesized that exposure to a greater number of MTP feedback cycles would lead to gains in student outcomes that are at least partially accounted for by teacher-student interaction.

2. Method

The present study predated the NCRECE investigation of MTP and was the first randomized controlled trial evaluation of MTP, at the time a newly designed professional development intervention (Pianta et al., 2008b). Pre-K teachers were recruited in the spring of the year prior to implementation through a process involving district-level pre-K program coordinators in the selected districts. All participating teachers and classrooms were a part of a state-funded pre-K program within a single state that served children who met "at-risk" status determined by state program criteria that included poverty, homelessness, parents with limited education, or family stress.

2.1. Recruitment procedures

School district, teacher, and child participants were recruited in several steps. School districts with classrooms participating in the state-funded pre-K program were selected for recruitment based on having more than one pre-K classroom; many districts in the state supported only one pre-K classroom and therefore were not recruited. Invitation letters were then sent to district-level coordinators that described the study, the intervention, and to ask about their interest in participating. Forty-one district coordinators agreed to facilitate recruitment of teachers, including three of the four largest districts in the state and more than one-third of all the districts in which the state-funded program was operating. District coordinators provided the study team with contact information for pre-K program teachers in their district and they co-signed a letter of recruitment indicating the district's permission for teachers to enroll in the study. Teachers received an individual letter inviting them to participate in the study.

The 41 participating districts were randomly assigned to one of the study conditions (i.e., MTP Coaching, Control), with the intent that all teachers within a district participate in the same study condition for the two consecutive years of the study. These assignments by district were first stratified by district size (large, medium, and small) and then assigned randomly to condition.

2.1.1. Teachers

District liaisons in the 41 districts that had agreed to participate in recruitment invited all eligible pre-K teachers in their programs to information and recruitment sessions. Teachers were eligible for participation if they were the lead teacher in a classroom in which the majority of children were eligible for kindergarten the following school year and did not have an IEP at the start of the current school year. Classroom instruction had to be conducted in English for most of the school day, and high-speed internet access at the program site was required. As noted above, each of the participating districts was randomly assigned to one of the two study conditions; all teachers within a district participated in the same study condition with the goal to participate across two consecutive academic years.

Across both years in which teachers were recruited and participated, 192 teachers provided data included in the analyses reported herein. A total of 173 teachers completed the full two years of the study, with 82 of those having been assigned randomly to MTP coaching and 91 to the control group. All teachers remained in their assigned group for the two years of the study. All available data from the 192 teachers and corresponding students from their classrooms were used in analyses. In the main evaluation study of MTP from which this dataset is drawn, the control group condition allowed teachers access to a set of literacy and language instructional activities and a website with informational resources on teacher-student interaction and a library of 1–2 min annotated video clips that depicted "high quality" interactions for each of the CLASS dimensions. The treatment and control teachers were compared on an assortment of teacher characteristics and no differences were detected on factors such as education, training, or race/ethnicity (Pianta et al., 2008b).

Ninety-five percent of the teachers were women. Most teachers reported their race/ethnicity as Caucasian (72%); 24% reported African American and 4% reported multi-racial. In terms of educational background, 66% had a BA degree and 35% had advanced degrees; 85% were specifically certified to teach 4-year-old children. Teachers had an average of 16 years of classroom experience, with a range of 1–43 years.

To estimate potential bias related to differential participation in MTP coaching as a function of teacher characteristics, analyses compared teacher and classroom characteristics for teachers who completed the minimum number of MTP cycles specified at the outset of the program (eight) and those who completed fewer. Analyses contrasted the MTP teachers who fully participated (eight coaching cycles) with the remaining teachers having a lower level of participation (see Mashburn et al., 2010). There were no statistically significant differences (all tests $p > .05$; all effect sizes $d < 0.04$) between the two groups of teachers regarding the percentage who had an advanced degree, the percentage with training in the field of early childhood education, or the number of years of experience teaching pre-K. In addition, there were no statistically significant differences in the classroom composition of teachers who did and did not fully participate regarding the percentages of children who were poor, mean maternal education levels of children in the classrooms, or children's mean pretest scores for four language or literacy assessments.

There were significant differences in classroom compositions for teachers who did and did not fully participate regarding the percentage of children with limited English proficiency ($F = -2.51$, $p < .05$) and the mean pretest scores on the blending sounds subtest ($F = 2.94$, $p < .01$). Effect sizes for these significant differences were small ($d < 0.05$). Specifically, as compared to classrooms of teachers with lower levels of participation, the fully participating teachers had a higher proportion of children enrolled in their classrooms with Limited English Proficiency (LEP) and lower pretest scores on the blending subtest.

2.1.2. Students

Teachers distributed child recruitment packets to parents of all children in their classroom (if they taught a.m. and p.m. sessions, they were asked to only send to the a.m. families); these packets included a letter explaining the project, a parental consent document, a Family Contact form, and a Family Questionnaire. Parents returned packets to the school. During fall of each academic year, four children among those whose parents provided consent to allow their child to participate were selected randomly from each classroom for the purposes of assessing child outcomes. The selection criteria included selecting two boys and two girls, when possible; when not possible four students were selected from the list of consented students. Among participating classrooms, complete child outcome assessment data were available for 1947 children (two cohorts enrolled for each of two years in which their teacher was participating in the study). The sample of children was ethnically diverse (i.e., 48% African American, 16% Latinx) with average levels of maternal education at just over a high school degree. Approximately 69% of preschool children included in the study were at or below the 1.50 poverty threshold, 55% were girls, and 20% spoke a language other than English at home.

Classrooms across conditions had similar proportions of boys and poor children. However, there were significant differences between conditions in the number of students per classroom (higher in the control condition) and the percent of LEP children (higher in MTP condition). To adjust for these differences, a full complement of teacher- and classroom-level variables were included as covariates in the models predicting outcomes from condition.

2.2. Measures

Measures include those collected on teacher-student interaction, child outcomes, teacher and child characteristics, and features of MTP.

2.2.1. Teacher–student interactions

All teachers submitted digital videos of their implementation of instructional activities in literacy, which were collected approximately every 2 weeks throughout the school year (September to June) during the period of MTP coaching. The nature and quality of teachers' instructional interactions on each video was rated using the 10 dimensions of interaction described by the CLASS observational measure (Pianta et al., 2008b). Factor analysis of the CLASS yields three factors (Hamre et al., 2012; Hamre et al., 2014), including (a) *Emotional Support*, (b) *Classroom Organization*, and (c) *Instructional Support*. Each dimension is described in a 7-point rating scale with behavioral indicators and anchor point descriptions provided for low, medium, and high levels of that dimension. Observers' ratings of teacher-child interactions on the 10 dimensions, from a minimum of four occasions across the school year, were the teacher outcome measure of focus in the present study.

A rigorous procedure for establishing and maintaining inter-rated agreement and to coding of observations was followed. Tapes of teachers' classroom instruction were coded by a team of observers trained to agreement and blind to condition. Training consisted of exposure to short video clips illustrating each CLASS dimension, followed by practice coding of at least five master coded videos of preschool classrooms. Discussions of these practice videos were led by a certified CLASS trainer. Following training, each observer had to pass a test in which they needed to score within 1 point of the master code on 80% of scores, across five additional video segments. If observers did not pass this test, additional training segments were provided and a second test was administered. These training procedures and criteria for inter-rater agreement are similar to those used in other studies using classroom observation methodology (e.g., National Institute of Child Health and Human Development Early Child Care Research Network, 2002; Pianta et al., 2005).

After passing the inter-rater agreement test, observers were assigned randomly to code tapes, and the coding team met on a weekly basis to conduct joint drift tests and discuss any coding-related concerns. Inter-rater agreement for each of the CLASS dimensions was computed by comparing the ratings made by two independent observers of 33 randomly selected tapes. Ratings that were within 1 point of each other along the 1–7 rating scale were considered to reflect an acceptable degree of accuracy. For the present study, the percentage of ratings that were within 1 point of each other for each of the CLASS dimensions included Positive Climate (85%), Teacher Sensitivity (82%), Instructional Learning Formats (79%), Behavior Management (94%), Productivity (97%), Concept Development (85%), and Language Modeling (85%). These levels of agreement are comparable to the inter-rater agreement data reported for the other scales of the CLASS that have been used in large scale observational studies of preschool through third grade classrooms that used live observation procedures (La Paro et al., 2004; National Institute of Child Health and Human Development Early Child Care Research Network, 2002, 2005). Studies have demonstrated that increasing the number of segments observed with CLASS reduces error associated with occasions and observers (Carbonneau et al., 2020; Mantzicopolous et al., 2018); in the present study, the scoring of segments drawn from four separate observation occasions is greater than what is typically reported from use of CLASS. Moreover, the random assignment of taped segments and the regular scoring of taped segments by the team of coders is consistent with efforts to reduce observer bias and increase agreement. Ratings for each CLASS dimension were averaged for each teacher across the observations and then composited to the three domains supported by factor analysis, including Emotional Support, Organizational Support, and Instructional Support.

2.2.2. Children's positive engagement

Teachers rated each students' skills on subscales from the Teacher-Child Rating Scale (Hightower, 1986), a 5-point Likert scale (1 = *not at all*, 5 = *very well*) assessing positive qualities of classroom adjustment as well as behavioral and learning problems. Task Orientation (e.g., functions well even with distractions; α's = 0.92–0.93); Social Skills (e.g., gets along well with others; α's = 0.93–0.94) and Peer Assertiveness (e.g., friendly toward peers; α's = 0.90–0.92) were used to describe students' positive adjustment in the classroom, consistent with Hightower (1986) in which these subscales load on a "positive adjustment" factor.

Teachers also completed an adapted version of the *Social-Conversational Skills Rating Scale* (SCSRS; Girolametto, 1996), a two-part checklist comprising 25 statements responded to using a 5-point Likert-type scale. This instrument examines children's skills during conversational exchanges, particularly their communicative assertiveness and responsiveness, which reflect two areas of skill that cannot be examined using formal test formats (e.g., standardized testing). The SCSRS was originally designed as a parent report measure for use in identifying preschool children's communicative profiles during dyadic contexts (Girolametto, 1996). Cronbach's alpha (0.85, 0.91 for responsiveness and assertiveness respectively), concurrent validity with MLU ($r = 0.55$), and test-retest reliability ($r = 0.90$ and 0.86 for responsiveness and assertiveness, respectively) have been documented with parents as respondents. Using teachers as informants, Cronbach's alpha for the total scale score was calculated as 0.91.

In the present study we used factor scores from the results of the sample-specific factor analysis to form a composite score reflecting teachers' reports of students *Positive Engagement in the Classroom.*

2.2.3. Children's early literacy outcomes

For children's language and literacy skills, we used the individually administered direct assessment, the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP; Lonigan et al., 2002). The Pre-CTOPPP was administered at the beginning of the fall semester (baseline measures) and near the end of the spring semester (outcome measures). It was designed for use with children from 3 to 5 years of age and is a precursor to the Test of Preschool Early Literacy (Lonigan et al., 2007). The Pre-CTOPPP provides scores from four subtests, including Blending, Elision, Print Awareness, and Receptive Vocabulary. Based on consultation with the assessment's authors, raw scores are reported and were used in analyses.

The *Blending* subtest includes items that measure whether children can blend initial phonemes onto one-syllable words, initial syllables onto two-syllable words, and ending phonemes onto one-syllable words. The *Elision* subtest measures whether children can break apart initial and ending phonemes, as well as initial syllables, from one- and two-syllable words. *Print Awareness* items measure whether children recognize individual letters and letter-sound correspondences, and whether they differentiate words in print from pictures and other symbols. *Receptive Vocabulary* items measure children's word knowledge. The Pre-CTOPPP subtests have shown adequate internal consistency, test-retest reliability, and concurrent validity in past research by the test developers and in several large studies, including the Head Start Impact Study, Institute for Education Sciences (IES) Even Start Classroom Literacy Interventions and Outcomes Study, IES Preschool Curriculum Evaluation Research Study, and IES Early Reading First National Evaluation (Lonigan et al., 2004; National Center for Education Evaluation (NCEE), 2007).

Teachers in this study administered the Pre-CTOPPP and PALS (see below) after receiving training by the investigators. Small-scale pilot tests have demonstrated the reliability and validity of data collected by assessments administered by teachers in Head Start classrooms (Invernizzi et al., 2004). In addition, most teachers had administered a similar literacy/language assessment to children in their classrooms in the past and were therefore very familiar with standardized testing procedures. At the beginning of the project, all teachers completed training focused on administration of the language and literacy battery, and fidelity of administration was randomly checked via videotape for 20% of teachers each fall. Teachers accurately administered standardized items over 90% of the time and reported that for 96% of the assessments children's performances were "most typical" or "very typical" of their usual classroom functioning.

Phonological Awareness Literacy Screening (PALS; Invernizzi et al., 2004) is a statewide individually administered assessment of early literacy. The PALS-PreK is a standardized, criterion-referenced assessment of children's emergent literacy skills that includes measurement of phonological awareness (Rhyme and Beginning Sounds subtests) and print knowledge (Alphabet Knowledge, Print Awareness subtests). It is designed for use with children who are 3–5 years of age. Raw scores for each of the individual subtests were used in analyses in the present study as an index of children's emergent literacy skills. Psychometric qualities include acceptable levels of test-retest and inter-rater reliability, and concurrent validity (Invernizzi et al., 2004). The PALS shows high test–retest, internal consistency reliability and correlates with concurrent and future individually administered reading assessments (see Invernizzi et al., 2004).

2.2.4. Classroom, teacher, and child characteristics

Parents of each child in the study completed a brief family questionnaire that assessed the child and family characteristics of gender, language spoken in the home, and years of maternal education. In addition, household income and the number of children and adults living in the household were used to calculate an income-to-needs ratio. A family was defined as "in poverty" using 150% of the Federal poverty guidelines as the threshold for the income-to-needs ratio. In the spring, teachers reported on the number of days each child was absent.

Data from family questionnaires completed by parents of all consented children (not just selected children) were used to calculate a *classroom-level measure of poverty*. This was computed as the proportion of children (ranging from 0 to 1) within each class who met this criterion. Across the classrooms, an average of 75% of parents completed the family questionnaire. Teachers reported on the number of children in their class, as well as the number of boys and children of LEP.

2.2.5. MTP coaching cycles

MTP coaching is implemented over the course of the academic year, with cycles taking approximately two weeks to implement from the start of one cycle to the start of the next cycle. With the wide range of circumstances across a school year (e.g., vacations, weather delays) it was expected there would be variation in the number of cycles and length of cycles. Given those realities, MTP coaches were trained to facilitate teachers' engagement in the MTP cycle to complete cycles in roughly a two-week period and to conduct as many cycles as possible over the course of the year, with eight as the target for a minimal level. Over the course of the year variation was present in the number of actual cycles completed by teachers. Each teacher in the MTP treatment group completed between 3 and 20 cycles, with an average of between 8 and 9 cycles.

2.3. Preliminary analyses: outcome factor analysis

Exploratory Structural Equation Modeling (ESEM, Asparouhov & Muthén, 2009) was used to evaluate the factor structure of two sets of measures: the TCRS for potential socioemotional skill constructs, and the Pre-CTOPPP and PALS literacy assessments for language and literacy constructs. ESEM allows cross-loadings of all variables onto the factors, but still allows the user to specify the number of factors, rotate loadings, and obtain valid goodness-of-fit statistics. All factor analyses were done at the individual student level, but separately for the fall and spring. Student nesting within teachers was accounted for by clustered standard errors. Loadings were estimated with Geomin rotation, which is valid for estimation structures with fewer than three factors and simple to moderate variable complexity (Asparouhov & Muthén, 2009).

A single factor best represented the TCRS subscales for social skills, peer assertiveness, and task orientation. These items had strong

Table 1
Exploratory structural equation model standardized loadings of socioemotional measures.*, **

	Positive engagement factor			
	Fall		Spring	
Measure	β	SE	β	SE
Social communication	0.53***	0.01	0.49***	0.01
Social skills	0.67***	0.02	0.66***	0.03
Peer assertiveness	0.70***	0.02	0.71***	0.03
Task orientation	0.85***	0.02	0.94***	0.02

Note. A one factor solution fit these measures best for both fall (CFI ≈ 1.00, TLI ≈ 0.96, RMSEA = 0.03, SRMR = 0.01) and spring (CFI ≈ 1.00, TLI ≈ 1.00, RMSEA = 0.03, SRMR = 0.01). See Method section for a description of Exploratory Structural Equation Modeling as a method of factor analysis.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

omega reliability (see Catalán, 2019) in both the fall ($\omega = 0.88$) and spring ($\omega = 0.89$), with all items positively loading onto the factor at both timepoints. Higher scores for the factor seemed to represent a child who is willing to engage with social events and classroom tasks with little inhibition; hence we refer to this factor as Positive Engagement in the Classroom. The fall model ($\chi^2(2) = 6.23$, $p = .04$) had strong fit (Comparative Fit Index [CFI] ≈ 1.00, Tucker-Lewis Index [TLI] ≈ 0.96, Root Mean Square Error [RMSEA] = 0.03, Standardized Root Mean Square [SRMR] = 0.01) based on the cutoffs suggested by Bandalos and Finney (2019). In spring, the model ($\chi^2(2) = 4.02$, $p = .13$) also fit well (CFI ≈ 1.00, TLI ≈ 1.00, RMSEA = 0.03, SRMR = 0.01). Table 1 includes the Positive Engagement factor loadings.

The PALS and pre-CTOPPP scales were explained by two factors. The first factor had strong loadings from Receptive Vocabulary, Blending, and Elision from the pre-CTOPP, as well as Rhyme Recognition from the PALS. This factor reflected children's overall *Phonological and Language* skills. Significantly loading items for this factor had adequate reliability in both the fall ($\omega = 0.79$) and spring ($\omega = 0.84$). The second factor was more confined, with the highest loadings from print knowledge (from both pre-CTOPP and PALS) and so was described as *Print Knowledge*. Significantly loading items for the Print Knowledge factor showed acceptable omega reliabilities in both the fall ($\omega = 0.71$) and spring ($\omega = 0.81$). For the fall, the ESEM factor structure ($\chi^2(19) = 85.34$, $p < .001$) had adequate fit indices (CFI = 0.97, TLI = 0.95, RMSEA = 0.05, SRMR = 0.02). For the spring, the item loadings were similar, and the model ($\chi^2(19) = 85.34$, $p < .001$) showed strong fit as well (CFI = 0.98, TLI = 0.96, RMSEA = 0.05, SRMR = 0.02). Table 2 contains the factor loadings for the Phonological and Language Skills factor and the Print Knowledge factor. These factors were merged into the student-level data as factor scores, and then aggregated to the classroom level for analysis.

2.4. Data analysis plan

Path modeling was used to assess the question of indirect effects of MTP: Do more cycles of the MTP intervention relate to higher end-of-year student outcomes through improved teacher-child interaction quality (i.e., Emotional Support, Classroom Organization, and Instructional Support)? The data consisted of variables for classroom composition (e.g., percentage of the class that was male), student outcomes, and teacher measurements (i.e., cycles, CLASS domain scores) that varied within a teacher by year of the study. We utilized clustered standard errors to account for this clustering of observations within teachers. Furthermore, considering sampling and intervention assignment was done from a pool of 41 districts, we accounted for stratification by district in our model's standard error estimates. Finally, to better account for variables with potentially non-normal distributions (see Table 3), we included robust standard error estimates.

The mediation model structure had the number of MTP cycles directly predict the spring factor outcome scores, as well as the three domains of spring teacher-student interaction mediators. These CLASS domain mediators also predicted the outcome factor scores. Both the spring factor outcomes and CLASS score mediators had respective fall scores as autoregressive covariates (e.g., fall Print Knowledge predicted spring Print Knowledge, fall Instructional Support predicted spring Instructional Support). Finally, four classroom composition covariates were accounted for in the mediator and outcome regressions, including the (a) number of students in the class, (b) percentage of students in the class that were White, (c) percentage of students that were male, and (d) percentage that were LEP. This model setup allows the estimation of both direct and indirect effects.

We also evaluated the extent to which significant direct and indirect paths detected in the first stage of analyses (i.e., number of cycles predicting CLASS domains) varied based on the teacher's fall CLASS domain scores. Therefore, fall CLASS scores were included as moderators in a separate, subsequent model evaluating the possibility that the number of intervention cycles demonstrated differential relations with spring CLASS domains depending on the teacher's fall score in that domain. For example, if the number of

Table 2
Exploratory structural equation model standardized loadings of language and literacy.

	Phonological & language skills factor				Print knowledge factor			
	Fall		Spring		Fall		Spring	
Measures	β	SE	β	SE	β	SE	β	SE
Language and literacy scale	−0.01	0.01	−0.09	0.07	0.74***	0.03	0.81***	0.07
Receptive vocabulary	0.68***	0.07	0.71***	0.08	0.00	0.06	−0.03	0.09
Blending	0.53***	0.07	0.56***	0.06	−0.04	0.07	0.04	0.06
Elision	0.72***	0.06	0.73***	0.02	0.00	0.05	0.00	0.01
Print awareness	0.18*	0.09	0.02	0.03	0.64***	0.08	0.79***	0.04
Print and word awareness	0.30**	0.10	0.22*	0.09	0.40***	0.10	0.51***	0.09
Beginning sound awareness	0.33***	0.10	0.26**	0.09	0.29**	0.09	0.43***	0.09
Rhyme awareness	0.36***	0.08	0.45***	0.10	0.27**	0.09	0.23*	0.11
Nursery rhyme awareness	0.47***	0.10	0.56***	0.11	0.10	0.11	0.05	0.12

Note. A 2-factor solution fit these measures best for both fall (CFI = 0.97, TLI = 0.95, RMSEA = 0.05, SRMR = 0.02) and spring (CFI = 0.98, TLI = 0.96, RMSEA = 0.05, SRMR = 0.02). See Method section for a description of Exploratory Structural Equation Modeling as a method of factor analysis.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 3
Classroom level variable descriptive statistics.

Classroom level variables		Mean	SD	Min	Max	Skewness	Kurtosis
# MTP cycles		5.52	6.20	0.00	19.00	0.54	−1.30
Emotional support	Fall	4.68	0.93	2.00	7.00	−0.15	−0.34
	Spring	4.53	0.94	1.67	6.50	−0.20	−0.46
Classroom organization	Fall	5.12	0.78	2.33	7.00	−0.55	0.79
	Spring	5.20	0.83	2.50	7.00	−0.40	−0.11
Instructional support	Fall	2.80	0.78	1.33	6.33	0.83	1.36
	Spring	3.02	0.77	1.00	5.67	0.14	0.30
Positive engagement factor	Fall	0.02	0.58	−1.49	1.83	0.06	−0.22
	Spring	0.03	0.56	−1.83	1.61	−0.12	0.19
Phonological and language skills factor	Fall	0.01	0.52	−1.30	1.37	0.44	−0.08
	Spring	−0.01	0.56	−1.72	1.25	−0.40	0.07
Print knowledge factor	Fall	0.00	0.51	−1.30	1.37	0.44	−0.08
	Spring	0.00	0.56	−1.77	1.38	−0.33	0.02
# Students in classroom [a]		14.51	2.74	5.00	19.00	−1.31	1.23
% White		0.27	0.29	0.00	1.00	1.03	0.00
% Male		0.50	0.12	0.10	0.88	−0.05	0.11
% Limited English proficiency		0.14	0.26	0.00	1.00	2.08	3.15

Note. N (Classrooms) = 301, n (Teachers) = 192. Factor scores were calculated at the student level and aggregated for each year up to the classroom level. Fall scores were aggregated from observations from August through September, while spring scores encompassed April through May.
[a] The number of students in the class at the beginning of the school year.

intervention cycles significantly predicted spring Emotional Support, then fall Emotional Support was included as an interactive term with cycles in a moderation model. This moderation term can have an indirect effect on student outcomes through the mediators (teacher-student interaction), which is tested in a manner identical to that used for other indirect effects (see below). A significant moderated mediation coefficient would suggest that potential indirect effects of MTP exposure on student outcomes can be stronger or weaker depending on the teacher's initial CLASS domain level.

Model estimation and analyses were conducted in Mplus version 8.1 (Muthén & Muthén, 2017).

2.4.1. *Evaluating model fit and indirect effects*

We used several criteria for considering the model as supporting mediation. First, the mediation model had to have strong fit based on the Bandalos and Finney (2019) cutoffs, with a CFI and TLI above 0.95, a RMSEA below 0.05, and a SRMR below 0.08. Second, the mediation model had to have been a better fit to the data than one without mediation paths (i.e., cycles only predicted the outcome factors, not the CLASS domains), as shown by a significant Satorra-Bentler Chi-squared difference test (Bryant & Satorra, 2012; Satorra & Bentler, 2001) between the mediation and non-mediation model. Third, at least one direct regression coefficient for cycles predicting CLASS domains had to be significant at $p < .05$, as well as a coefficient for the CLASS domain predicting the factor outcomes. Finally, the indirect effect (i.e., the product of multiplying the significant direct coefficients) had to have a 95% CI that did not contain zero. These criteria also applied for testing our moderated mediation hypothesis (see above).

When testing indirect effects, because they are the product of two coefficients, the distribution of coefficients is non-normal (Bollen & Stine, 1990; MacKinnon et al., 2004) and must be estimated in other ways to validly determine if the product coefficient is significantly different from zero. We utilized Selig and Preacher's (2008) Monte Carlo integration method for estimating 95% CIs in non-normal indirect effect distributions, which is based on the magnitude, variance, and covariance of the coefficients. CIs that did not include zero were considered significant and indicated a high likelihood that student outcomes were impacted by the number of cycles of coaching through the intervention's effect on teacher-student interaction quality (i.e., CLASS domain scores).

2.4.1.1. Inclusion of control teachers. Intervention and control teachers were differentiated by whether they engaged in any MTP cycles or not. In other words, control teachers had zero cycles, whereas intervention teachers had one or more cycles. This approach of assigning control teachers a value of zero on the number of cycles did not skew the distribution of cycles (see skewness and kurtosis values in Table 3) and we employed estimation methods that were robust to violations of non-normality (i.e., ESTIMATOR = MLR in MPlus). Assigning zeros to the control teachers has several beneficial effects, including increasing the variation within the number of cycles to highlight the relations of the cycles to the teacher performance mediators. The full range of teachers and students can also be used in this case, providing more valid distributions for CLASS domain scores, covariates, and student outcomes. A larger sample size was especially helpful in utilizing structural equation modeling, which requires moderately high sample sizes to ensure accurate standard error estimates.

2.4.1.2. Missing data and model sample. To retain all available data in our models, we used the model-based imputation technique of Full Information Maximum Likelihood (FIML; see Enders & Bandalos, 2001, for an explanation of FIML, and for evidence from a simulation study that supported its use above other missing data methods). FIML requires all continuous, exogenous variables to covary, which does not affect regression coefficient estimates. Furthermore, using the full range of teachers and students in the data provides a larger sample size, along with more valid distribution and standard error estimates for variables. After retaining missing

Table 4
Classroom-level pair-wise correlations.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. # MTP Cycles	–																
2. ES Fall	0.19*	–															
3. ES Spring	0.28*	0.24*	–														
4. CO Fall	−0.05	0.51*	0.19*	–													
5. CO Spring	−0.01	0.21*	0.61*	0.34*	–												
6. IS Fall	0.17*	0.52*	0.09	0.40*	0.10	–											
7. IS Spring	0.17*	0.12	0.66*	0.23*	0.53*	0.18*	–										
8. Engagement Fall	0.22*	0.10	0.14*	0.01	−0.04	0.04	0.00	–									
9. Engagement Spring	0.21*	0.15*	0.20*	0.01	−0.03	0.03	0.08	0.65*	–								
10. PLS Fall	0.00	0.02	0.03	0.07	−0.04	0.03	−0.02	0.37*	0.28*	–							
11. PLS Spring	−0.04	0.05	0.05	0.19*	0.04	0.11	0.01	0.18*	0.30*	0.65*	–						
12. PK Fall	0.10	0.07	0.08	−0.01	−0.04	0.01	0.01	0.52*	0.37*	0.81*	0.47*	–					
13. PK Spring	0.02	0.07	0.09	0.10	0.04	0.08	0.06	0.26*	0.44*	0.56*	0.85*	0.57*	–				
14. # Students	−0.20*	−0.15*	−0.22*	−0.04	−0.13	−0.04	−0.15*	−0.19*	−0.09	0.08	0.11	−0.01	0.08	–			
15. % White	0.02	0.05	0.09	0.03	0.03	−0.02	0.07	0.16*	0.10	0.25*	0.12*	0.13*	−0.01	−0.06	–		
16. % Male	−0.03	0.03	−0.15*	0.01	−0.06	0.04	−0.09	−0.06	−0.01	−0.01	0.02	−0.05	0.00	0.03	−0.02	–	
17. % Limited English Proficiency	0.42*	0.06	0.18*	−0.04	0.01	0.05	0.23*	0.13*	0.13*	−0.16*	−0.26*	0.05	−0.11	−0.24*	−0.36*	0.03	–

Note. ES is Emotional Support, CO is Classroom Organization, IS Instructional Support, PLS is Phonological & Language Skills, and PK is Print Knowledge.
* $p < .05$.

Table 5
Standardized regression coefficients for covariates in mediation model.

	Spring endogenous variable					
Covariate	Emotional support	Classroom organization	Instructional support	Engagement factor	Phonological & language skill factor	Print knowledge factor
# Students in Classroom	−0.11 (0.06)	−0.09 (0.06)	−0.05 (0.07)	0.05 (0.04)	0.03 (0.06)	0.05 (0.06)
% White	0.14* (0.07)	0.03 (0.09)	0.19* (0.09)	0.01 (0.06)	−0.11* (0.05)	−0.16** (0.05)
% Male	−0.13* (0.06)	−0.07 (0.06)	−0.09 (0.06)	0.04 (0.04)	0.04 (0.04)	0.04 (0.05)
% LEP	0.13* (0.07)	0.02 (0.07)	0.26*** (0.07)	0.02 (0.06)	−0.26*** (0.05)	−0.23*** (0.06)
Fall Autoregression	0.19*** (0.05)	0.25*** (0.06)	0.20*** (0.06)	0.63*** (0.04)	0.57*** (0.04)	0.56*** (0.04)

Note. These covariates were included in the mediation model regressions (see Fig. 1) for the three CLASS domains and the three student outcome factors. Fall autoregression refers to the same construct as the spring variable, with fall variables containing measures from August through September, whereas spring variables encompassed April through May.

* $p < .05$.
** $p < .01$.
*** $p < .001$.

data, we utilized 301 data-points between both years of the study, drawn from 192 teachers.

3. Results

3.1. Descriptive statistics

Table 3 contains the list of descriptive statistics for all the variables used in the analyses and Table 4 provides the correlations between classroom-level variables. As noted in the tables, all values were in the expected range.

3.2. Overall mediation model fit

The full mediation model, that included covariates (child, teachers, and classroom characteristics) as well as fall CLASS scores along with the focal variables of MTP cycles, teacher-student interactions, and outcomes, showed high fit indices (CFI = 0.99, TLI = 0.97), a low error (RMSEA = 0.04), and a low residual (SRMR = 0.03), all of which surpassed cutoffs for good model fit. Furthermore, the mediation model showed better fit as compared to a model that lacked the mediation regression from cycles to the CLASS domain mediators (*Satorra-Bentler* $\Delta\chi^2 = 14.79$, $\Delta df = 3$, $p = .002$), showing that the stronger model for the data accounts for mediation and indirect effects.

3.3. Coefficients and indirect effects

In the mediation model, more intervention cycles for a teacher were related to higher Emotional Support in the spring ($b = 0.02$, $\beta = 0.15$, $p = .03$), adjusting for fall Emotional Support and the covariates (i.e., number of students in the classroom, percentage of students in the class that were white, percentage that were male, and percentage that were LEP). Higher spring Emotional Support was in turn related to higher average Positive Engagement Style in the spring ($b = 0.12$, $\beta = 0.21$, $p = .005$), again accounting for fall Positive Engagement and the covariates. Fig. 1 provides a depiction of this model and Table 5 provides the covariate regression coefficients.

This mediation pathway (Cycles → Spring Emotional Support → Spring Positive Engagement) was tested using a Monte Carlo indirect effect test (Preacher & Hayes, 2008), and the product of unstandardized coefficients was significantly different from zero (95% CI [0.0001185, 0.006634]). This primary result suggests that when teachers were exposed to a greater number of MTP cycles, the resulting improvements in their Emotional Support interactions over the course of the intervention were associated with increases in ratings of students' Positive Engagement behaviors for the children in their classroom.

No significant indirect effects were detected for the analyses examining the path models for the students' literacy and language outcomes.

3.4. Moderated mediation

Based on the mediation model, the moderated-mediation model tested if the direct relation between the number of MTP cycles and spring Emotional Support was dependent on the teacher's fall-level Emotional Support for a given year (i.e., cycles interacted with fall Emotional Support), and if so whether this moderation also affected the indirect effect of cycles influencing Approach Style through spring Emotional Support. However, the interaction term did not show significance, which suggests that all teachers, regardless of initial Emotional Support, can equally benefit from more cycles of the intervention.

4. Discussion

4.1. Summary of findings

Coaching takes a wide variety of forms and has become a primary focus for professional development of teachers (Hemmeter et al., 2015) in efforts to improve their classroom practices (e.g., Diamond & Powell, 2011; Domitrovich et al., 2009; Downer et al., 2012; Pianta et al., 2008b) and student learning (Clements et al., 2013; Hamre et al., 2012). Recent reviews (e.g., Snyder et al., 2015) have synthesized the fundamental aspects of coaching and have identified common elements of observation/data collection and feedback on observed practice that repeat in cyclical form across a period of time. The present study examined the fundamental question of the extent to which teachers' improvements in the quality of their interactions were acquired as a function of exposure to a specific coaching model (Pianta et al., 2008b) and which were in turn related to student outcomes. In other words, the study examined the extent to which the mediational process presumed to underlie coaching models—that coaching improves teachers' interactions in the classroom which in turn improve student outcomes—can be confirmed empirically. In this context the study focuses specifically on variation in the number of coaching cycles in which teachers engaged as an indicator of dosage. The results have implications for further research on questions concerning minimal dosage levels, or thresholds, and for questions concerning costs relative to benefits for this form of professional development (Snyder et al., 2015).

4.2. Implications

The results of the present study provide limited support for the premise that coaching can improve student outcomes as a function of improvements it induces in teachers' behaviors with students. In the present set of analyses, when teachers were exposed to a greater number of coaching cycles focused on their classroom interactions, ratings of students' positive classroom engagement with teachers, peers, and tasks increased because of associations with the increased quality of teachers' emotionally supportive interactions with students. These findings generally confirmed a similar investigation of MTP indirect effects conducted in a separate sample (Pianta et al., 2021) that found a significant indirect effects for coaching dosage on student literacy outcomes as a function of mediation by teachers' instructional behavior. This indirect linkage of coaching dosage, teacher-student interaction, and student outcomes confirmed the hypothesized effects associated with MTP, and of coaching more generally. It is also clear that more research is needed to better specify the nature of these associations and the specific elements of coaching that are responsible for the effects detected.

As presented earlier, a single cycle of MTP contains several elements, and in this way the focus on dosage of cycle masks these more specific features. Like dosage, some of features of coaching cycles in MTP (feedback in an area of practice) are also present in other coaching models (Clements et al., 2013; Domitrovich et al., 2009; Hemmeter et al., 2015; Wasik & Hindman, 2011) that have shown impacts on teacher practice and on child outcomes and are present in general frameworks for coaching (e.g., Snyder et al., 2015). For example, Snyder et al. describe sessions (a dosage index similar to cycles), observation (or in our case exposure to video), and feedback (focus of prompts) as common features of coaching and present coaching impacts on child outcomes as mediated by effects on teacher practice. The present study is one of a few to have established empirical support for the mediation processes that most of these models have in common, which is dosage of cycles. As we noted earlier, more research is needed to further identify the specific elements of coaching inputs that may be contributing to these effects.

It is not a surprise that the pathway from MTP coaching cycles to improved aspects of students' positive engagement in the classroom involves improvements in teachers' emotional supportive interactions. In several recent studies involving observations of pre-K classrooms using CLASS or other approaches, the emotional responsiveness of teachers is predictive of not only teacher reports of students' social adjustment and teacher-student relationships, but also direct assessments of students working memory and inhibitory control (e.g., Pianta et al., 2020). Furthermore, investigations have indicated that students' positive engagement in the classroom, or skills such as inhibitory control, mediate gains in academic skills (Raver et al., 2008; Wasik & Hindman, 2011).

The scientific evidence for the possible benefits of coaching, whether paired with a specific curriculum (Bierman et al., 2008; Clements et al., 2013; Domitrovich et al., 2009; Hsieh et al., 2009; Raver et al., 2008; Wasik & Hindman, 2011) or not (Downer et al., 2012; Pianta et al., 2008b), along with the current study's support of the underlying logic of its impact, would seem to support policies and investments in its widespread use (e.g., Lloyd & Modlin, 2012). However, it is also the case that beyond broad descriptive efforts to guide programmatic investments in coaching as a widespread intervention (Barton et al., 2011; Shannon et al., 2015), there are considerable gaps in knowledge about implementation supports that help ensure investments in coaching yield the expected or desired returns. For example, nearly all the scientific evidence in support of the benefits of coaching described in this paper and in the other investigation of MTP (Pianta et al., 2021) comes from highly structured, focused approaches in which the interventions are designed to target specific teacher behaviors and specific child outcomes related to those behaviors and implemented under well-resourced conditions by highly trained and supervised practitioners. Conversely, when states or school districts invest in coaching, these investments typically are not tied to a specific model of coaching and the target coaching actions, expected teacher behaviors, and desired student outcomes are often not highly specified in advance; in this sense the coaching approach is not designed as an aligned and integrated model (Snyder et al., 2015). In these cases, training and implementation supports often are weak or diffuse, leading to increased likelihood of weak or null effects. The present study suggests that at a minimum it is important to focus on ensuring teachers are steadily engaged to receive higher dosage levels of coaching support.

Regarding the coaching model used in the present study, there have been a few instances of large-scale implementation of MTP in early education programs, with results suggesting modest impacts on teacher behavior (e.g., Early et al., 2017) and on student outcomes (Dallas Independent School District, 2019). Given the findings of the present study, such impacts on teacher behavior may presage promising impacts on student outcomes, although at this point those evaluations have yet to be conducted. However, even in these implementations, there is considerable effort to provide robust implementation support to ensure sufficient dosage of exposure; this support may exceed that for a typical coaching intervention or approach.

4.3. Limitations and suggestions for future research

The present study has several limitations. Working across program locations or sites, in diverse program structures, and with varying curricula being implemented in classrooms (see LoCasale-Crouch et al., 2012) prevented the use of student outcome assessments more directly aligned to the curriculum in use. Such aligned student assessments may be more sensitive to improvements in teacher-student interaction, a possibility suggested by impacts of MTP on high school students' scores on state standards assessments (Allen et al., 2011). The lack of alignment and differences in literacy curricula and activities across program locations and sites may in part have accounted for the lack of significant findings for student literacy outcomes. The assessments of student outcomes in the present study, although standard in the field, were more generalized rather than aligned to classroom- or program-level activities or curricula. Additionally, although the sample was large and diverse, several teachers from the original sample dropped out of the study, the vast majority due to life or work circumstances (e.g., a family moving to a new location, teachers reassigned to new grade levels, a decision to leave teaching), with few resulting from not wishing to participate. Attrition in this study, despite being a considerable challenge methodologically and operationally, reflects the instability and churn of the early education workforce (Ryan & Whitebook, 2012), which likewise confronts broader efforts to improve the quality and impact of the early education workforce.

5. Conclusions

In sum, we found evidence to support a fundamental premise of nearly all models used for coaching teachers—that targeted feedback drawn from observation of their classroom practices can promote improvements in those practices, which in turn leads to gains in children's learning (Snyder et al., 2012). More specifically, we found that variation in dosage of MTP coaching (web-mediated, structured feedback, and analysis of teacher-student interaction using CLASS as the lens) may foster teachers' overall responsiveness to children's cues and their facilitation of students' positive engagement in the classroom, both in terms of social adjustment with teachers and peers and orientation to tasks. Within this sample of teachers, these coaching experiences were offered across a wide range of program sites, each varying in demographic, curricular, and program features. These results have implications for policies and investments in early education quality improvement efforts, particularly those that focus on teacher professional development and coaching, suggesting the importance of focused and structured feedback aligned to a standard measurement of practice.

The present study is just one effort to identify the components of coaching and its implementation that are associated with changes in student and teacher outcomes (Sheridan et al., 2009) based on the prominence of coaching (of all forms and rigor) in efforts to improve the quality and impact of educational opportunities. Further research is needed to examine the specific elements of coaching; in the case of MTP, this might involve teachers' exposures to videos of their practice, qualities of the teacher-coach relationship, or elements of the written feedback on their interactions that teachers receive. There is some evidence to suggest that different domains of teachers' interactions (e.g., emotional support or instructional support) are uniquely sensitive to feedback in that specific domain (and not other domains) and change at different rates across the year (Pianta et al., 2014). Given the promising results associated with coaching and its frequency of use in teacher professional development, systematic and programmatic empirical research on the processes and implementation of coaching may yield valuable insights.

References

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*(6045), 1034–1037. https://doi.org/10.1126/science.1207998

Ansari, A., & Pianta, R. C. (2018). Effects of an early childhood educator coaching intervention on preschoolers: The role of classroom age composition. *Early Childhood Research Quarterly, 44*, 101–113. https://doi.org/10.1016/j.ecresq.2018.03.001

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*(3), 397–438.

Bandalos, D. L., & Finney, S. J. (2019). Factor analysis: Exploratory and confirmatory. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (2nd ed., pp. 98–122). Routledge.

Barton, E. E., Kinder, K., Casey, A. M., & Artman, K. M. (2011). Finding your feedback fit: Strategies for designing and delivering performance feedback systems. *Young Exceptional Children, 14*, 29–46. https://doi.org/10.1177/1096250610395459

Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., Blair, C., Nelson, K., & Gill, S. (2008). Promoting academic and social-emotional school readiness: The head start REDI program. *Child Development, 79*, 1802–1817. https://doi.org/10.1111/j.1467-8624.2008.01227.x

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, 115–140.

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling, 19*(3), 372–398. https://doi.org/10.1080/10705511.2012.687671

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*, 579–620. https://doi.org/10.4236/jss.2017.58012

Carbonneau, K., Van Orman, D., Lemberger-Ruelove, M., & Atencio, D. (2020). Leveraging the power of observations: Locating sources of error in the individualized classroom assessment scoring system. *Early Education and Development, 31*, 84–99.

Catalán, H. E. N. (2019). Reliability, population classification and weighting in multidimensional poverty measurement: A Monte Carlo study. *Social Indicators Research, 142*(3), 887–910. https://doi.org/10.1007/s11205-018-1950-z

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45*(2), 443–494. https://doi.org/10.3102/0002831207312908

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies: Persistence of effects in the third year. *American Educational Research Journal, 50*(4), 812–850. https://doi.org/10.3102/0002831212469270

Dallas Independent School District. (2019). *Quality classrooms lead to better STAAR scores*. Interim report. Dallas Independent School District.

Desimone, L. (2009). How can we best measure teacher's professional development and its effects on teachers and students? *Educational Researcher, 38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Diamond, K. E., & Powell, D. R. (2011). An iterative approach to the development of a professional development intervention for Head Start teachers. *Journal of Early Intervention, 33*, 75–93. https://doi.org/10.1177/1053815111400416

Domitrovich, C. E., Gest, S. D., Gill, S., Bierman, K. L., Welsh, J. A., & Jones, D. J. (2009). Fostering high quality teaching with an enriched curriculum and professional development: Head start REDI. *American Educational Research Journal, 46*, 567–597. https://doi.org/10.3102/0002831208328089

Downer, J. T., Lopez, M. L., Grimm, K., Hamagami, A., Pianta, R. C., & Howes, C. (2012). Observations of teacher-child interactions in classrooms serving Latinos and dual language learners: Applicability of the classroom assessment scoring system in diverse settings. *Early Childhood Research Quarterly, 27*, 21–32. https://doi.org/10.1016/j.ecresq.2011.07.005

Downer, J. T., Pianta, R. C., Fan, X., Hamre, B. K., Mashburn, A., & Justice, L. (2011). Effects of web-mediated teacher professional development on the language and literacy skills of children enrolled in prekindergarten programs. *NHSA Dialog: Research-to-Practice Journal for the Early Childhood Field, 14*(4), 189–212.

Early, D. M., Maxwell, K. L., & Ponder, D. (2017). Improving teacher-child interactions: A randomized control trial of making the most of classrooms and my teaching partner professional development models. *Early Childhood Research Quarterly, 38*, 1–14. https://doi.org/10.1016/j.ecresq.2016.08.005

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Farran, D., Lipsey, M., & Wilson, S. (2011). *Experimental evaluation of the tools of the mind pre-K curriculum*. Peabody Research Institute, Vanderbilt University.

Fukkink, R. G. (2007). Does training matter? A meta-analysis and review of caregiver training studies. *Early Childhood Research Quarterly, 22*, 294–311. https://doi.org/10.1016/j.ecresq.2007.04.005

Girolametto, L. (1996). *Social-conversational skills rating scale*. Author.

Hamre, B. K., Hatfield, B. E., Jamil, F., & Pianta, R. C. (2014). Evidence for general and domain-specific elements of teacher-child interactions: Associations with preschool children's development. *Child Development, 85*(3), 1257–1274. https://doi.org/10.1111/cdev.12184

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2012). Promoting young children's social competence through the preschool PATHS curriculum and MyTeachingPartner professional development resources. *Early Education and Development, 23*(6), 809–832. https://doi.org/10.1080/10409289.2011.607360

Hemmeter, M. L., Hardy, J. K., Schnitz, A. G., Adams, J. M., & Kinder, K. A. (2015). Effects of training and coaching with performance feedback on teachers' use of pyramid model practices. *Topics in Early Childhood Special Education, 35*(3), 144–156. https://doi.org/10.1177/0271121415594924

Herren, J. K. (2009). *Being an effective mentor-coach (head start bulletin 80: Mental health)*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Hightower, A. D. (1986). The Teacher–Child Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review, 15*, 393–409.

Hsieh, W., Hemmeter, M. L., McCollum, J. A., & Ostrosky, M. M. (2009). Using coaching to increase preschool teachers' use of emergent literacy teaching strategies. *Early Childhood Research Quarterly, 24*, 229–247. https://doi.org/10.1016/j.ecresq.2009.03.007

Invernizzi, M., Sullivan, A., Swank, L., & Meier, J. (2004). *PALS pre-K: Phonological awareness literacy screening for preschoolers* (2nd ed.). University Printing Services.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *Elementary School Journal, 104*(5), 409–426. https://doi.org/10.1086/499760

Landry, S. H., Anthony, J. L., Swank, P. R., & Monseque-Bailey, P. (2009). Effectiveness of comprehensive professional development for teachers of at-risk preschoolers. *Journal of Educational Psychology, 101*, 448–465. https://doi.org/10.1037/a0016746

Li, H., Liu, J., & Hunter, C. V. (2019). A meta-analysis of the factor structure of the Classroom Assessment Scoring System (CLASS). *The Journal of Experimental Education, 10*, 1–24.

Lloyd, C. M., & Modlin, E. L. (2012). *Coaching as a key component in teachers' professional development: Improving classroom practices in head start settings*. OPRE report 2012-4. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from http://www.acf.hhs.gov/sites/default/files/opre/coaching_key.pdf.

LoCasale-Crouch, J., DeCoster, J., Cabell, S., Hamre, B., Downer, J., & Pianta, R. (2012). *Variation in teachers' instructional interactions within two interventions: Associations with intervention responsiveness and teacher/classroom characteristics* (Research No. ED530499). Society for Research on Educational Effectiveness. Retrieved from http://eric.ed.gov/PDFS/ED530499/pdfve.

Lonigan, C. J., McDowell, K., & Phillips, B. (2004). Assessment of children's pre-literacy skills. In B. Wasik (Ed.), *Handbook on family literacy: Research and services* (pp. 525–550). Lawrence Erlbaum.

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *Test of preschool early literacy*. Pro-Ed.

Lonigan, C. J., Wagner, R. K., Torgeson, J. K., & Rashotte, C. A. (2002). *Preschool comprehensive test of phonological & print processing*. Authors.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4

Mantzicopoulos, P. Y., French, B. F., Patrick, H., Watson, J. S., & Ahn, I. (2018). The stability of kindergarten teachers' effectiveness: A generalizability study comparing the framework for teaching and the classroom assessment scoring system. *Educational Assessment, 23*, 24–46.

Mashburn, A. J., Downer, J. T., Hamre, B. K., Justice, L. M., & Pianta, R. C. (2010). Consultation for teachers and children's language and literacy development during pre-kindergarten. *Applied Developmental Science, 14*(4), 179–196. https://doi.org/10.1080/10888691.2010.516187

Muthén, L. K., & Muthén, B. O. (2017). *Mplus version 8.1 [computer software]*. Muthén e& Muthén.

National Center for Education Evaluation (NCEE). (2007). *National evaluation of Early Reading first: Final report to congress*. U.S. Department of Education.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *Elementary School Journal, 102*(5), 367–387. https://doi.org/10.1086/499709

National Institute of Child Health and Human Development Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *The Elementary School Journal, 105*, 305–323. https://doi.org/10.1086/428746

Pianta, R., Hamre, B., Downer, J., Burchinal, M., Williford, A., LoCasale-Crouch, J., … Scott-Little, C. (2017). Early childhood professional development: Coaching and coursework effects on indicators of children's school readiness. *Early Education and Development, 28*(8), 956–975. https://doi.org/10.1080/10409289.2017.1319783

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*, 144–159. https://doi.org/10.1207/s1532480xads0903_2

Pianta, R. C., DeCoster, J., Cabell, S., Burchinal, M., Hamre, B. K., Downer, J., … Howes, C. (2014). Dose–response relations between preschool teachers' exposure to components of professional development and increases in quality of their interactions with children. *Early Childhood Research Quarterly, 29*(4), 499–508. https://doi.org/10.1016/j.ecresq.2014.06.001

Pianta, R. C., La Paro, K., & Hamre, B. K. (2008a). *Classroom assessment scoring system (CLASS)*. Paul H. Brookes.

Pianta, R. C., Mashburn, A., Downer, J., Hamre, B., & Justice, L. (2008b). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*(4), 431–451. https://doi.org/10.1016/j.ecresq.2008.02.001

Pianta, R. C., Whittaker, J. E., Vitiello, V., Ruzek, E., Ansari, A., Hofkens, T., & DeCoster, J. (2020). Children's school readiness skills across the pre-K year: Associations with teacher-student interactions, teacher practices, and exposure to academic content. *Journal of Applied Developmental Psychology, 66*. https://doi.org/10.1016/j.appdev.2019.101084

Pianta, Robert, Lipscomb, Daniel, & Ruzek, Erik (2021). Coaching teachers to improve students' school readiness skills: Indirect effecrs of teacher-student interaction. *Child Development, 92*(6), 2509–2528. https://doi.org/10.1111/cev.13600

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development on head start teachers and children. *Journal of Educational Psychology, 102*(2), 299–312. https://doi.org/10.1037/a0017763

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891.

Raver, C. C., Jones, S. M., Li-Grining, C. P., Metzger, M., Smallwood, K., & Sardin, L. (2008). Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in head start settings. *Early Childhood Research Quarterly, 23*(1), 10–26. https://doi.org/10.1016/j.ecresq.2007.09.001

Ryan, S., & Whitebook, M. (2012). More than teachers: The early care and education workforce. In R. C. Pianta (Ed.), *Handbook of Early childhood education* (pp. 92–110). Guilford Press.

Sabol, T. J., Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can rating pre-K programs predict children's learning? *Science, 341*(7148), 845–846. https://doi.org/10.1126/science.1233517

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507–514. https://doi.org/10.1007/BF02296192

Selig, J. P., & Preacher, K. J. (2008, June). *Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects* [Computer software]. Available from http://quantpsy.org/.

Shannon, D., Snyder, P., & McLaughlin, T. (2015). Preschool teachers' insights about web-based self-coaching versus on-site expert coaching. *Professional Development in Education, 41*(2), 290–309. https://doi.org/10.1080/19415257.2014.986819

Sheridan, S. M., Edwards, C. P., Marvin, C. A., & Knoche, L. L. (2009). Professional development in early childhood programs: Process issues and research needs. *Early Education & Development, 20*, 377–401. https://doi.org/10.1080/1040928080258795

Snyder, P., Hemmeter, M. L., Artman, K., Kinder, K., Pasia, C., & McLaughlin, T. (2012). Characterizing key features of the early childhood professional development literature. *Infants and Young Children, 25*, 188–212. https://doi.org/10.1097/IYC.0b013e31825a1ebf

Snyder, P., Hemmeter, M. L., & Fox, L. (2015). Supporting implementation of evidence-informed practices through practice-based coaching. *Topics in Early Childhood Special Education, 35*(3), 133–143. https://doi.org/10.1177/0271121415594925

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

Williford, A. P., & Shelton, T. L. (2008). Using mental health consultation to decrease disruptive behaviors in preschoolers: Adapting an empirically supported intervention. *Journal of Child Psychology and Psychiatry, 49*(2), 191–200. https://doi.org/10.1111/j.1469-7610.2007.01839.x