

Effects of Training and Coaching With Performance Feedback on Teachers' Use of *Pyramid Model* Practices

Topics in Early Childhood Special Education
2015, Vol. 35(3) 144–156
© Hammill Institute on Disabilities 2015
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0271121415594924
tecse.sagepub.com



Mary Louise Hemmeter, PhD¹, Jessica K. Hardy, PhD¹,
Alana G. Schnitz, MEd, BCBA¹, Jessie Morris Adams, PhD¹,
and Kiersten A. Kinder, PhD¹

Abstract

Training and coaching with performance feedback has been effective for supporting teachers to use evidence-based instructional practices. However, coaching with performance feedback has primarily been used to support teachers to use discrete skills, and there has been little evidence of maintenance and generalization. The purpose of this study was to examine the effects of a professional development intervention on teachers' implementation of practices related to the *Pyramid Model for Promoting Social-Emotional Competence in Young Children*, as well as the extent to which teachers generalized and maintained those practices. A multiple probe design across sets of *Pyramid Model* practices replicated across three teachers was used in this study. All teachers acquired the practices and maintained the practices after coaching ended. There was some evidence of generalization for all three teachers. The effects of teacher implementation on classroom-wide incidences of challenging behavior were mixed. Teachers all rated the coaching positively.

Keywords

environment, classroom, social, development, social skills, intervention strategies, teachers, personnel, single-subject designs, experimental studies, research methodologies

Recent reviews of the literature indicate that the prevalence rates of challenging behavior in preschool settings are between 14% and 34% (Qi & Kaiser, 2003; Upshur, Wenz-Gross, & Reed, 2009). Without intervention, children with challenging behavior may have difficulty interacting with their peers and may fall behind in academic skills (Gilliam, 2005). Challenging behavior can lead to children being expelled from preschool programs, further isolating them from effective interventions and peers with whom to interact. Gilliam (2005) conducted a survey of educators in 40 states and found the preschool expulsion rate to be approximately 6.67 per 1,000 preschool children enrolled. This rate is 3.20 times higher than the expulsion rate of students in K-12th grades (Gilliam, 2005).

Given these data, it is clear that early childhood educators need support around addressing the needs of children with ongoing challenging behavior. Early childhood educators have consistently identified addressing the needs of children with challenging behavior as one of the major challenges of their job and thus as a primary training need (Friedman-Krauss, Raver, Neuspiel, & Kinsel, 2014; Snell et al., 2012). Furthermore, teachers who report having access to behavior support personnel are less likely to ask for children to be removed from their classrooms (Gilliam,

2005). The provision of training for early childhood educators and access to behavior supports is consistent with implementation science, which suggests that it is the combination of effective intervention practices and effective implementation practices that lead to positive outcomes for children and families (Fixsen, Blase, Duda, Naoom, & Van Dyke, 2010).

Intervention Practices: *Pyramid Model* for Addressing Social-Emotional Competence

The *Pyramid Model for Promoting Social-Emotional Competence in Young Children* (Fox, Dunlap, Hemmeter, Joseph, & Strain, 2003) is a comprehensive, tiered model for promoting young children's social-emotional development and addressing challenging behavior. The *Pyramid Model*

¹Vanderbilt University, Nashville, TN, USA

Corresponding Author:

Jessica K. Hardy, Peabody College, Vanderbilt University, 110 Magnolia Circle, One Magnolia Circle Bldg., Rm 318C, Nashville, TN 37203, USA.
Email: jessica.k.hardy@gmail.com

includes evidence-based practices designed to support young children's social-emotional skills, and prevent and address challenging behavior in the classroom. The *Pyramid Model* includes three tiers of support. The first tier includes practices that are implemented universally and are related to creating high-quality, supportive environments and nurturing and responsive relationships. Some practices associated with this tier of the *Pyramid Model* are having meaningful conversations with children, supporting children through play, designing engaging learning centers and activities, implementing a predictable schedule, teaching behavior expectations, using positive descriptive feedback, and planning transitions. The second level of the *Pyramid Model* involves providing targeted social-emotional supports, including teaching friendship skills, emotional literacy, anger management, and social problem-solving. For children who need more intensive support around social-emotional skills, teachers implement individualized instruction with sufficient intensity for each child. The tertiary tier of the *Pyramid Model* involves the use of targeted interventions for individual children whose challenging behavior or other social-emotional needs are not responsive to developmentally appropriate guidance procedures. Targeted interventions are developed by a team, involve functional assessment, and include three components: prevention strategies, teaching new skills, and responding in ways that support the child's acquisition of those new skills. The effectiveness of the *Pyramid Model* in supporting all children in an inclusive setting depends on the concurrent implementation of practices across all tiers.

Implementation Supports: Training and Ongoing Coaching

For interventions, such as the *Pyramid Model*, to be implemented with fidelity, implementation supports, including ongoing training and coaching, are needed (Fixsen et al., 2010). A review of the professional development literature identified coaching with performance feedback as a successful intervention to change teacher behavior (Snyder et al., 2012). Performance feedback involves the use of information from observations of the teacher to provide specific feedback on the teacher's use of targeted teaching practices. The impact of coaching with performance feedback on teachers' use of specific social-emotional teaching and behavior support practices in preschool classrooms has been examined in several published studies. In these studies, researchers have examined the effects of coaching with performance feedback on a variety of discrete practices associated with the *Pyramid Model*, including praise and precorrections (Barton, Pribble, & Chen, 2014; Fullerton, Conroy, & Correa, 2009; Hemmeter, Snyder, Kinder, & Artman, 2011; Stormont, Smith, & Lewis, 2007).

In addition, a series of studies have examined the use of training followed by coaching with performance feedback on teachers' use of practices across tiers of the *Pyramid Model*. Fox, Hemmeter, Snyder, Binder, and Clarke (2011) used a multiple probe across skills design replicated across teachers to investigate whether coaching with performance feedback was effective in increasing teachers' use of *Pyramid Model* practices. The teachers were trained to use strategies associated with each tier of the *Pyramid Model*. For all three teachers, there was a functional relation between coaching and increased implementation of *Pyramid Model* practices. Furthermore, there was some evidence of maintenance of skills following the end of intervention.

Hemmeter and colleagues (Hemmeter, Fox, & Snyder, 2013) conducted a cluster randomized control trial in 40 preschool classrooms in two Southern states across two school years. In this study, intervention teachers received a 3-day intensive training, followed by approximately 14 weeks of weekly in-class coaching, debriefing meetings, and email feedback. The results indicated statistically significant differences between control and intervention teachers in the use of *Pyramid Model* practices as measured by the Teaching Pyramid Observation Tool (TPOT; Fox, Hemmeter, & Snyder, 2014). There was also evidence that teachers' use of the *Pyramid Model* practices was related to teacher-reported improvements in social skills and challenging behavior as well as observed improvements in the social skills of children who had the most challenging behavior.

Two additional studies were conducted on the effects of training and coaching with performance feedback on teachers' use of *Pyramid Model* practices. In these studies, feedback was delivered through email. Artman-Meeker and Hemmeter (2013) examined the effects of coaching with performance feedback, delivered via email, on two teaching teams' use of transition preparations, rule reminders, and social-emotional teaching strategies in preschool classrooms using a multiple baseline across strategies design. The data showed a functional relation between coaching and both teams' use of transition preparations, rule reminders, and social-emotional teaching practices. The effects of the intervention on children's challenging behavior were mixed, as there was a functional relation between increases in the teachers' use of targeted practices and the target child's behavior for only one of the two teaching teams/classrooms.

Artman-Meeker, Hemmeter, and Snyder (2014) used a cluster randomized control design in 33 Head Start classrooms to examine the effects of training and distance coaching on teachers' use of *Pyramid Model* practices. In this study, teachers in both groups received a 1-day training on *Pyramid Model* practices, and intervention teachers received coaching with performance feedback. The observations occurred via video, and feedback was delivered via

email. The results showed some statistically significant differences between control and intervention teachers in classroom quality. Furthermore, non-experimental analyses suggested that intervention teachers who accessed the intervention (i.e., read their email feedback and viewed training videos) more often used more *Pyramid Model* practices compared with both control teachers and intervention teachers who accessed the intervention less frequently. There were also reductions in child challenging behavior in the classrooms of teachers who accessed the intervention more frequently.

Although previous studies provide support for the use of training and coaching with performance feedback on the use of *Pyramid Model* practices, additional research is needed to replicate previous results and address issues related to intensity of coaching, maintenance and generalization of the effects of coaching, and teacher report of satisfaction and feasibility of the intervention. The current study was designed to address these issues using a professional development intervention that included training and coaching with performance feedback. Specifically, the following research questions were addressed:

Research Question 1: Is training and coaching effective for increasing teachers' use of practices related to the *Pyramid Model*?

Research Question 2: Do teachers generalize the use of coached practices to activities other than those in which they were coached?

Research Question 3: Do teachers maintain practices after coaching on those practices end?

Research Question 4: Does implementing the *Pyramid Model* practices with fidelity decrease classroom-wide instances of challenging behavior?

Research Question 5: Does implementation of the *Pyramid Model* overall improve when teachers receive training and coaching on specific *Pyramid Model* practices?

Research Question 6: What are teachers' perspectives of the coaching process, coaching relationship, and sustainability of the *Pyramid Model* practices?

Method

Participants

Three teachers in an urban school district participated in the study. Teachers were recruited from the control group who participated in a randomized control trial conducted by the same research team the previous school year (Hemmeter et al., 2011). Teachers from the control group who had expressed an interest in coaching were contacted, and three teachers who consented were included in the study. During the randomized trial, control teachers received a 2-day

training on the *Pyramid Model*, approximately 6 months before the beginning of the current study, but did not receive any type of follow-up support.

Bianca was a Caucasian female with 8 years of teaching experience. She had a master's degree and certification in early childhood special education (ECSE) and K-12 modified special education. Kendra, a Caucasian female, held a bachelor's degree in ECSE and was working toward a master's degree. Her certification was also in ECSE. She had 2 years of teaching experience. Susan was a Caucasian female with a master's degree and certification in ECSE. She had 6 years of teaching experience.

Setting

The study took place in blended preschool classrooms in three elementary schools. All classrooms had between 14 and 16 children, about half who had disabilities, and all classrooms had a lead teacher and an assistant teacher. Each teacher had 2 to 4 children with persistent, ongoing challenging behavior and a high percentage of children receiving free or reduced price lunch (87.5%–93.8%). Observations took place in the classrooms during the regular school day. Coaching sessions took place in the classroom, during naptime or after school.

Materials

Several types of materials were used in the study. Checklists, described below, were the primary data collection tool. Personal digital assistants (PDAs) and the Multi-Option Observation System for Experimental Studies (MOOSES; Tapp, Wehby, & Ellis, 1995) were used to collect data on children's challenging behavior. Each live coaching session was recorded using handheld audio recorders to enable collection of fidelity data.

Measurement

Data were collected on (a) teachers' use of targeted *Pyramid Model* practices (checklists), (b) instances of classroom-wide challenging behavior (Class-Wide Challenging Behavior Observation Tool [CCBOT]), and (c) teachers' global use of all *Pyramid Model* practices (TPOT). Primary data were collected in the morning and included a large-group activity, centers/small groups, and transitions. In addition, generalization data were collected after lunch for each teacher.

The primary behaviors of interest were the teacher's use of specific practices associated with the *Pyramid Model*. These behaviors were measured through the use of researcher-designed checklists that were based on an earlier version (Hemmeter, Fox, & Snyder, 2008b) of the TPOT (Fox et al., 2014). Nine checklists were developed, and each

checklist contained 7 to 10 indicators related to the practice, with precise criteria for receiving credit for each indicator. For example, for the checklist related to teaching children classroom behavior expectations, specific indicators included the following: (a) teacher provides instructions or reminders on posted behavior expectations to individual children, during play, and within small-group activities and (b) throughout the observation, teacher provides specific positive feedback to children on meeting posted behavior expectations. The checklist score was a percentage, which was calculated by dividing the number of indicators present by the number of indicators possible and multiplying by 100. These data were collected approximately 1 to 2 times per week. During each observation, the coach collected data on the teacher's current set of targeted practices. In addition, the coach collected intermittent probe data on the other sets of targeted practices during at least 30% of data collection observations.

Children's challenging behavior was measured using the CCBOT (Hemmeter, Fox, & Snyder, 2008a). The purpose of the CCBOT was to provide a measure of the frequency of challenging behavior in classrooms; however, information on the reliability and validity of this measure is not available. Data were collected in two 45-min CCBOT observations that were conducted 4 times during the study: in baseline and after intervention on each set of targeted practices for each teacher. Data collectors used a partial interval recording system with 10-s intervals to record the presence or absence of both low-intensity and high-intensity challenging behaviors by scanning the classroom and noting whether any challenging behavior occurred in each interval. Low-intensity challenging behaviors were defined as behaviors that distract children or a teacher from typical activities, routines, or instructions. Examples of low-intensity challenging behavior were not following directions, taking a toy from another child, and wandering around the room not engaged with an activity or routine. High-intensity challenging behaviors were defined as behaviors that disrupt the flow of classroom activities and routines. They generally require adult intervention to prevent physical harm to people or materials/property or to continue with classroom routines and activities. Examples of high-intensity challenging behavior were property destruction, throwing objects, physical aggression with a peer, and tantrums.

An earlier version (Hemmeter et al., 2008b) of the TPOT (Fox et al., 2014) was used to measure global changes in teachers' fidelity of implementation of *Pyramid Model* practices over the course of the study. The TPOT is a measure of teachers' fidelity of implementation of *Pyramid Model* practices. It is administered by conducting an approximately 2-hr observation and 15- to 20-min interview with teachers. The version of the TPOT used in the present study had 108 indicators organized under 15 key *Pyramid Model* practice items. Indicators are scored either

yes (practice was observed or reported to be implemented during the interview) or *no* (practice was not observed or reported to be implemented during the interview). In addition to key practice items, 16 red flags and 7 environmental arrangement indicators were included on this version of the TPOT. Red flags are practices that are inconsistent with *Pyramid Model* practices. An example would be reprimanding children for expressing their emotions, which is incompatible with the key practice item related to teaching children to express emotions. The TPOT was used 4 times during the study: prior to baseline and after intervention on each set of targeted practices for each teacher. Teachers' performance on the TPOT was analyzed by calculating the percentage of environmental items present, the percentage of indicators present for each of 15 key practice items, and the percentage of red flags not present.

Although the TPOT and CCBOT data were collected by data collectors, checklist data were collected by the coaches. Checklist data and other information from the observations were necessary for the coach to effectively provide feedback. If the coach did not collect the primary data, it would have been necessary to schedule additional observations to allow the coach to gather information needed for coaching sessions. However, because the coach served as the primary data collector for her teacher, it was important to measure if the teacher's changes in behaviors related to the intervention were dependent on the coach's presence in the classroom. Thus, a different data collector observed and collected data periodically throughout the intervention phase for each targeted practice. The teacher was unaware of the purpose of these observations, and the coach was not present during these observations. These data will hereafter be referred to as alternate observer checks. In addition, interobserver agreement (IOA) data were collected on at least 33% of the observation sessions to ensure that the coach's data were reliable.

Experimental Design

A multiple probe design across sets of practices, replicated across teachers, was used to evaluate the effectiveness of coaching with performance feedback on teachers' use of *Pyramid Model* practices. In multiple probe designs, experimental control is demonstrated through the staggered introduction of the independent variable to different behaviors, with immediate changes in behavior only after introduction of the independent variable (Gast & Ledford, 2010). The intervention involved two phases for each practice: (a) baseline during which no feedback was provided by the coach and (b) intervention during which the coach used coaching and performance feedback targeted around the current practice until the teacher implemented the indicators at criterion levels. Each tier of the multiple probe design represented a different set of targeted practices.

General Procedures

After teachers consented to participate in the study, they were assigned a coach. The teachers and coaches met at a prestudy kick-off event, during which the study timeline and procedures were explained to teachers and the teachers and coaches met one another. After the study kick-off, TPOT data were collected. Then, each coach conducted an informal observation in her teacher's classroom to familiarize herself with the classroom. After those observations, each coach met with her teacher to set goals for coaching. During this meeting, the coach described the study and the coaching process and helped the teacher identify three goals to target during the study. Each potential goal was related to a *Pyramid Model* set of practices (e.g., providing balanced and predictable schedules and routines, teaching children classroom behavior expectations, promoting emotional literacy, and teaching social problem-solving) and a corresponding checklist. The coach used the pre-study TPOT data, as well as her own observation notes, to narrow the field of possible goals/checklists to ones that would most benefit the teacher and to help the teacher select action plan goals. The coach provided data from the TPOT observation to the teacher to guide the selection of action plan goals. The teacher was encouraged to choose goals that were foundational and in need of improvement (based on the TPOT data). In addition, the teacher was encouraged to choose a goal related to targeted social-emotional support that was of strong interest to the teacher. Bianca's goals were as follows: (a) schedules and routines (targeted practices Set 1), (b) behavior expectations (targeted practices Set 2), and (c) problem-solving (targeted practices Set 3). Kendra's goals were as follows: (a) behavior expectations (targeted practices Set 1), (b) schedules and routines (targeted practices Set 2), and (c) emotional literacy (targeted practices Set 3). Susan's goals were as follows: (a) behavior expectations (targeted practices Set 1), (b) schedules and routines (targeted practices Set 2), and (c) problem-solving (targeted practices Set 3). The order in which goals were addressed were based on whether the practices were foundational or targeted, as well as what order seemed most logical and appropriate, given the teacher's individual classroom and child needs. After determining the goals and the order in which to address them, baseline data collection began.

Baseline Procedures

Each teacher had a minimum of three baseline observations. Baseline observations took place during the target activities that had been determined during the goal setting meeting with the teacher. These typically included a teacher-directed large-group activity, a transition, and center time.

In baseline observations, the coach collected data on the three checklists that corresponded to the teacher's action

plan goals. The coach did not meet with the teacher or provide coaching during the period of time when the baseline observations were being conducted. At the end of the initial baseline observations, CCBOT data were collected in each classroom.

While intervention was being provided for the first set of targeted practices, baseline probe data were collected on the other sets of targeted practices. Probe data collection on the practices in which teachers were not currently being trained occurred once per week and in three consecutive observations before phase changes. Intervention in the first set of targeted practices occurred only after stable baseline data were observed. Intervention in the second set of targeted practices occurred only after the teacher reached the criterion in the first set of targeted practices and had stable baseline data for the second set of targeted practices. Intervention in the third set of targeted practices occurred only after the teacher reached the criterion in the second set of targeted practices and had stable baseline data for third set of targeted practices.

Intervention Procedures

After baseline data were stable, the intervention began. The intervention procedures used in this study were consistent with our previous work (e.g., Fox et al., 2011; Hemmeter et al., 2011). For each set of targeted practices, the intervention began with a session that included training and action plan refinement. Although goals were selected prior to baseline so they could be measured during baseline, specific action plans were not developed until the beginning of intervention for each set of target practices. The coach provided a short (30–60 min) training on the practices related to the action plan goal/checklist. Training included a standardized PowerPoint presentation, video examples, and discussions on how to individualize the practices to the teacher's classroom. The coach also provided four implementation guides, each focusing on *Pyramid Model* practices. The coach and teacher worked together to develop steps for meeting the teacher's first action plan goal. The coach and teacher decided on the order in which steps would be addressed, a timeline for implementing the steps, and materials needed to implement each step. As the teacher worked on action plan steps, the coach provided the teacher with implementation materials related to the teacher's action plan, such as problem-solving visuals and center choice systems. In addition to the initial trainings, it was sometimes necessary to provide "booster" trainings. These booster trainings occurred after school breaks of at least 2 weeks (due to scheduled holiday breaks and/or snow days). Booster trainings involved reviewing relevant content from the first training, as well as the teacher's action plan. The use of booster training is shown in Figures 1 to 3 where a "B" is present.

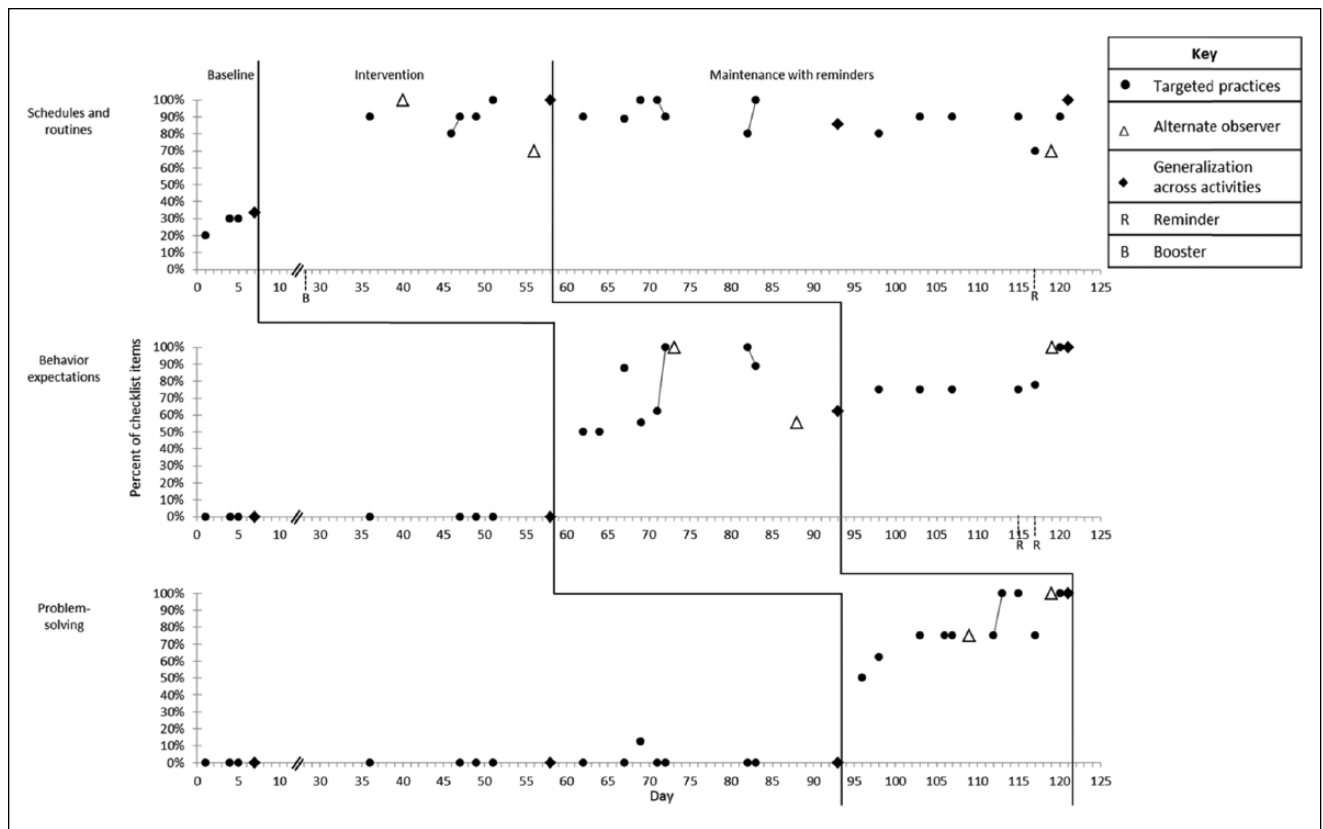


Figure 1. Bianca's acquisition, generalization, and maintenance of targeted practices.

After each training and action planning session, the coach conducted live coaching observations, debriefing meetings, and email feedback. In addition to the data collection observations described earlier, the coach conducted a coaching observation once a week without collecting specific data. The purpose of these observations was to provide direct coaching (e.g., prompting the teacher, providing descriptive praise following the teacher's use of practices, etc.). It was not possible to collect data and provide live coaching during the same observations.

All observations were followed by either live or email feedback. Feedback was delivered 3 times per week. Approximately 70% of observations were followed by live debriefing meetings (approximately twice per week), and approximately 30% of sessions were followed by email feedback (approximately once per week), except in weeks in which scheduling difficulties prohibited this. The debriefing meetings and email feedback were based on data and notes collected during the checklist data collection observation or during the live coaching observation. Debriefing meetings were individualized to meet the needs of the teacher, but they all followed the same general format, which is consistent with the protocol used in previous studies (e.g., Fox et al., 2011; Hemmeter et al., 2011). The

coach (a) checked in with the teacher to see whether there was anything she would like to share about her classroom, (b) reviewed the action plan goal, (c) shared data on the action plan goal, (d) provided supportive feedback, (e) provided constructive feedback, (f) provided additional resources, (g) discussed scheduling upcoming observations and meetings, and (h) asked the teacher if she had questions or concerns. Email feedback followed the same protocol, with adjustments to accommodate the differences between face-to-face and email communication (e.g., the inclusion of the graph of the teacher's data as an email attachment).

The teacher received coaching around her current action plan goal until she reached the criterion, which was 80% of checklist indicators for three consecutive data collection observations. The checklist data were shared with the teacher as part of the coaching process, and the teacher was aware of the criterion. When providing coaching to the teacher, the coach only provided feedback on the current action plan goal/set of practices. She did not provide instruction or coaching on topics relevant to future or previous action plan goals. If the teacher's data dropped below criterion on previously addressed action plan goals, the coach provided a reminder to continue using the previous

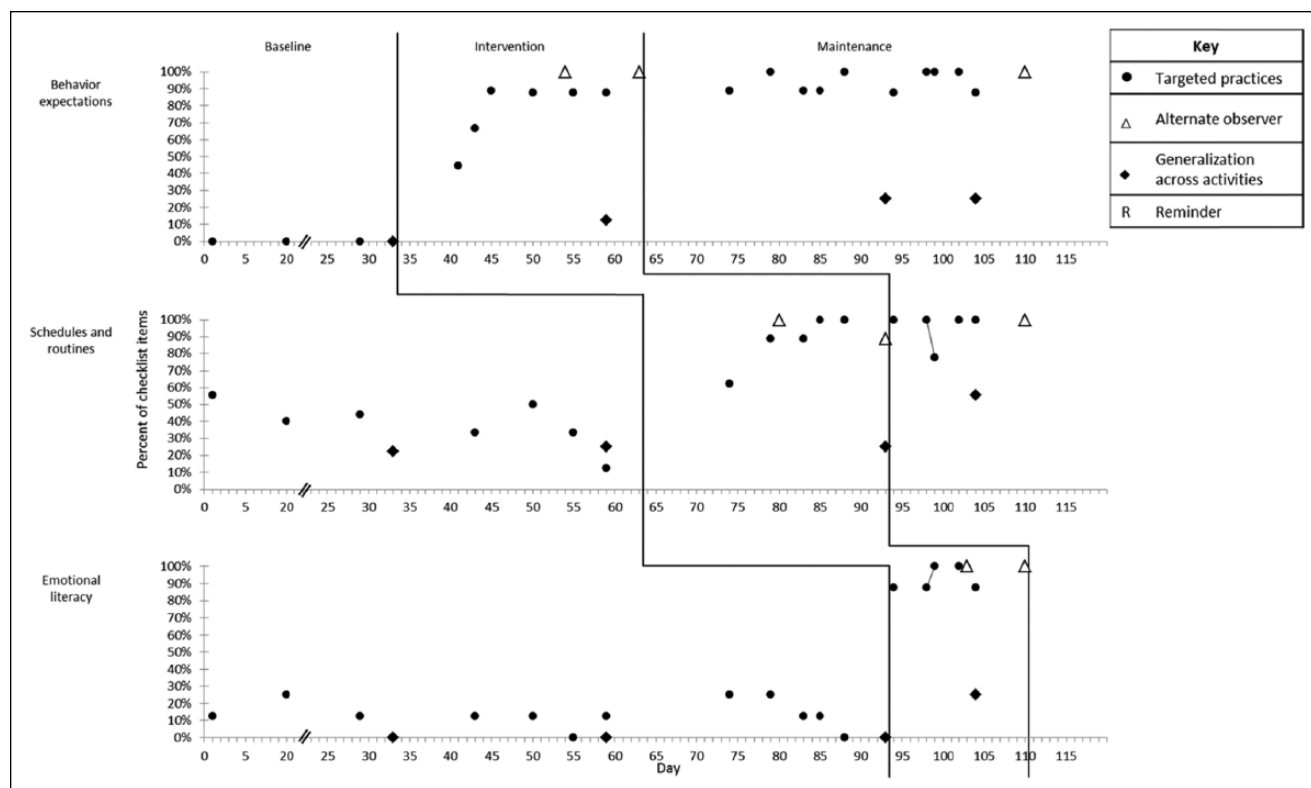


Figure 2. Kendra's acquisition, generalization, and maintenance of targeted practices.

practices but provided no other feedback. The coach documented the occurrence and content of these reminders.

A variety of strategies were used in coaching observations, debriefing meetings, and email feedback, such as (a) providing materials, (b) modeling, (c) helping in the classroom, (d) problem-solving, (e) reflective conversation, (f) environmental arrangement, (g) side-by-side verbal or gestural support, (h) goal setting and planning, and (i) graphing.

Generalization Procedures

Generalization across activities was measured periodically for all targeted practices and throughout all phases of the study using the checklists. The coach and teacher identified an additional time of the day for the coach to observe, which was afternoon in two classrooms (Kendra and Susan) and after lunch in the late morning in one classroom (Bianca). The teacher did not receive coaching about using the practices during this time of the day. Because of classroom scheduling issues (e.g., lunch, nap, playground), the generalization activities were often quite different than the target activities. For example, in one classroom, generalization across activities was measured during "quiet centers," which occurred directly after rest time, and "wrap-up circle." Although these activities were similar in name to the activities on which the teacher was coached

(i.e., centers and large group), in practice, they were quite different. In addition, some checklist indicators could not be evaluated during the generalization activities and were scored as "no opportunity." For example, one indicator in the schedules and routines checklist was the presence of both large- and small-group activities. In some classrooms, this was not possible due to the limitations of the schedule (i.e., small groups did not occur in the afternoon). The teacher was unaware of the purpose of the generalization observations.

Maintenance Procedures

After the teacher reached the criterion on an action plan goal, coaching was not delivered on that goal but data continued to be collected intermittently. During this time, the coach did not provide feedback on the action plan goal but could provide reminders about continuing to use the practices when there was a drop in the teacher's performance. These reminders differed from the booster sessions in that the reminders were simply a reminder "to continue to use the practices." The reminders did not include a review of the practices or additional information about the practices as was done in the booster sessions. The frequency and use of reminders varied across teachers and can be seen in Figures 1 to 3 where a "R" is present.

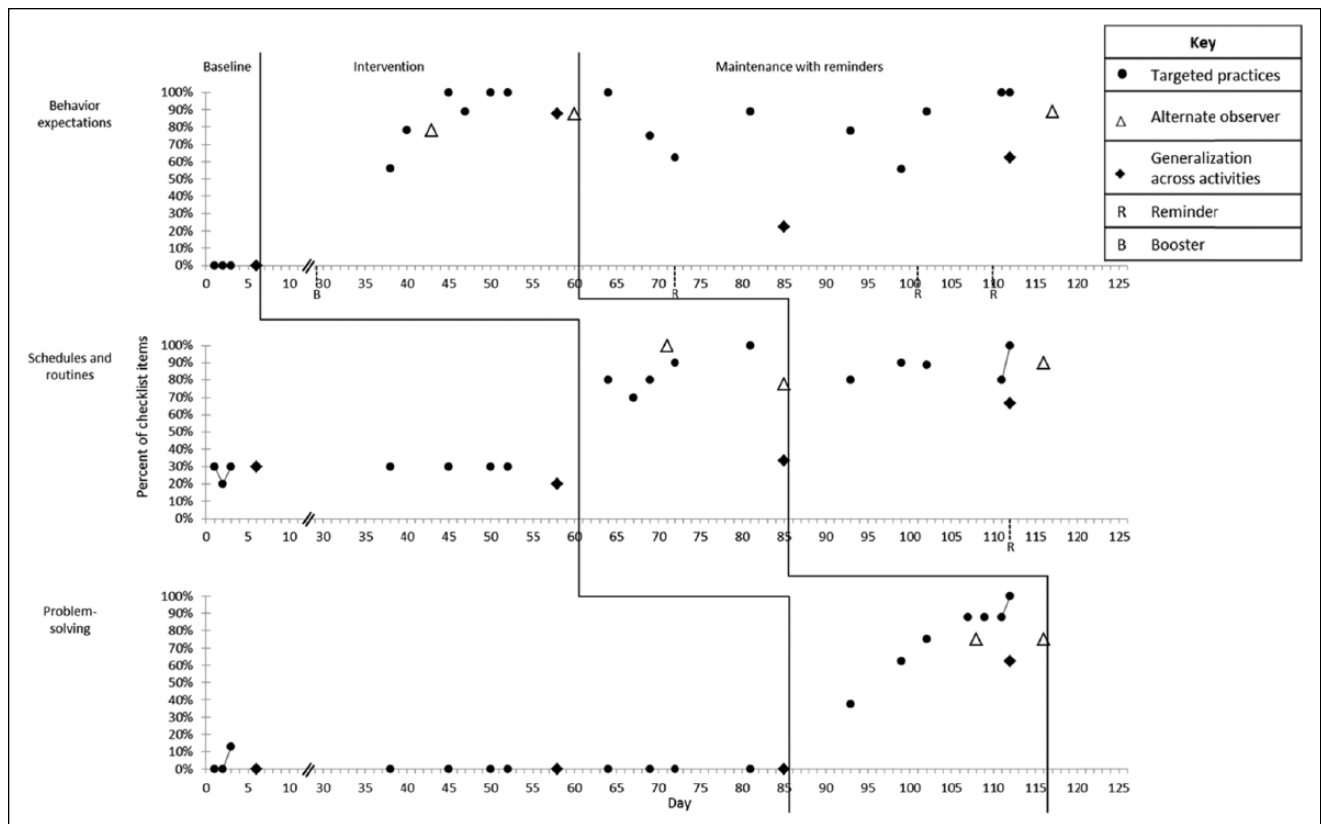


Figure 3. Susan's acquisition, generalization, and maintenance of targeted practices.

IOA

Before the study began, data collectors were trained on each tool and practiced using each tool in non-participating classrooms. They were required to be reliable on each tool prior to collecting data for the study. For teacher checklists, each data collector was required to complete two observations (paired with two different observers) at 80% reliability for each checklist to be considered reliable. To be considered reliable on classroom and child measures (i.e., TPOT and CCBOT), each data collector had to complete three observations with an already trained data collector, with at least 80% agreement on the measure being used.

During the study, IOA data were collected for all teacher, classroom, and child measures. At least 30% of observations using each measure were conducted with a primary and reliability data collector. The percentage agreement between the two data collectors was calculated using a point-by-point formula: The number of agreements divided by the number of agreements plus disagreements was multiplied by 100.

Procedural Fidelity

Procedural fidelity data were collected on at least 20% of each type of coaching session for each coach (i.e., goal

setting, training and action planning, debriefing, email, closing). All coaching sessions were audio recorded, and all coaching emails were saved. Coaching sessions and emails were randomly selected to be reviewed by a procedural fidelity data collector. The data collector used a checklist when listening to the audio recordings and viewing the emails to determine if the coach followed the protocol for each type of session. Examples of procedural fidelity items include reviewing the action plan goal, providing positive support feedback, and providing constructive feedback. Procedural fidelity percentages were calculated by dividing the number of items present by the number of items possible and multiplying by 100.

In addition, to ensure procedural fidelity was completed reliably, another data collector independently completed the procedural fidelity checklists for at least 20% of all sessions that were reviewed for procedural fidelity. IOA between the two procedural fidelity data collectors was calculated using point-by-point agreement.

Social Validity

At the conclusion of the study, each teacher completed a social validity questionnaire. Using a 6-point, Likert-type scale (1 = *strongly agree* to 6 = *strongly disagree*), teachers

rated the following: usefulness of components of the coaching framework, usefulness of various coaching strategies, adequacy of the dosage of coaching, relationship with coach, and sustainability of practices. Teachers were not required to put their name on the questionnaire and were given the option to return the completed surveys via U.S. mail to the first author.

Results

The results will be described as follows: (a) IOA, (b) procedural fidelity, (c) acquisition of targeted practices, (d) alternate observer checks, (e) generalization of targeted practices, (f) maintenance of targeted practices, (g) classroom-wide challenging behavior, (h) TPOT, and (i) social validity.

IOA

IOA was calculated for each data collection measure. IOA data were collected in 41.28% of checklist observations, 33.33% of generalization checklist observations, 41.67% of TPOT observations, and 41.67% of CCBOT observations. For the primary measure, the checklists, the mean IOA was 93.45%, with a range of 56% to 100%. For generalization sessions, IOA on the checklists had a mean of 97.15%, with a range of 80% to 100%. The mean reliability on the TPOT was 87.98%, with a range of 84.40% to 90.15%. Mean agreement on the CCBOT was 99.71%, with a range of 98.9% to 100%.

Procedural Fidelity

Procedural fidelity was collected on 44.90% of live coaching sessions and 40.74% of emails. Fidelity percentages were high across interventionists and session types. The mean fidelity for live coaching sessions was 99.43%, with a range of 90.9% to 100%. The mean fidelity for emails was 100%. Reliability on fidelity was measured in 31.88% of the sessions coded for fidelity, with 100% agreement.

Acquisition of Targeted Practices

The first three research questions relate to teachers' acquisition of targeted *Pyramid Model* practices, generalization of these practices to activities other than those in which they were coached, and maintenance of practices. Data from the *Pyramid Model* practice checklists were used to answer these questions. The metric used was percentage of indicators present on the checklist. Acquisition of the sets of targeted practices (i.e., percentage of checklist items for a goal) is indicated in Figures 1 to 3 by the closed circles.

Bianca's baseline data were low and stable for all three sets of practices. After coaching began on her use of

schedules and routines, there was a significant shift in level, as can be seen in Figure 1. The double hash marks in the abscissa of the graph indicate school breaks of more than 2 weeks. There was a delay between the end of baseline and intervention data collection because there was no school due to unanticipated snow days and winter break. No coaching or data collection was possible during this break. This break was followed by a booster training on the first set of targeted practices. After Bianca's intervention data for the first set of targeted practices were stable (at least three consecutive data points at or above the criterion of 80% of checklist indicators), coaching began on her second action set of targeted practices (i.e., behavior expectations). There was an immediate and significant shift in level for this set of practices. After intervention data were stable, coaching began on Bianca's third action set of targeted practices (i.e., problem-solving). There was an immediate and significant shift in level that stabilized before ending intervention.

Kendra's baseline data were low and stable for her first set of targeted practices (i.e., behavior expectations) as can be seen in Figure 2. After coaching began on this set of targeted practices, there was an immediate shift in the level of the data. After Kendra's intervention data for the first set of targeted practices were stable, coaching began on her second set of targeted practices (i.e., schedules and routines). Although baseline data for this set of targeted practices were variable, there was an immediate and significant shift in level for this set of practices. After these data were stable and at criterion, coaching began on her third set of targeted practices (i.e., emotional literacy), which resulted in an immediate and significant shift in level of the data.

Susan's baseline data were low and stable for all three sets of practices that were targeted for coaching, as can be seen in Figure 3. After coaching began on the first set of practices (i.e., behavior expectations), there was a significant shift in the level of the data. As with Bianca, this shift was delayed due to school breaks, which is indicated by the double hash marks in the abscissa of her graphs. This break was followed by a booster training on the first set of targeted practices. Once these data were stable, coaching began on her second set of targeted practices (i.e., schedules and routines). There was an immediate and significant change in level for this set of practices. A similar pattern was seen for her third set of targeted practices (i.e., problem-solving).

Alternate Observer Checks

Alternate observer checks were conducted to ensure that the teacher used targeted practices when her coach was not present and collecting data. A data collector other than the coach collected data during these observations. During the alternate observer checks, each teacher demonstrated practices she was being coached on or had already acquired even when her coach was not present at about the same

Table 1. TPOT and CCBOT Data.

Teacher	Time	TPOT data			CCBOT data		
		Percentage of indicators (%)	Number of environmental items	Number of red flags	Percentage of intervals with challenging behavior present		
					Total ^a (%)	Low intensity (%)	High intensity (%)
Bianca	Baseline	28.70	3	5	25.74	23.33	2.96
	After Set 1	43.52	4	0	16.67	12.04	4.81
	After Set 2	39.81	5	1	15.37	13.33	2.04
	After Set 3	49.07	7	1	10.56 ^b	10.37 ^b	0.19 ^b
Kendra	Baseline	41.67	5	2	24.26	17.59	9.44
	After Set 1	44.44	7	2	21.48	12.78	8.70
	After Set 2	58.33	7	0	1.11	0.93	0.19
	After Set 3	44.44	6	2	2.78	1.30	1.48
Susan	Baseline	34.26	6	5	4.44	0.74	3.70
	After Set 1	46.30	7	3	13.15	11.30	1.85
	After Set 2	38.89	7	4	9.81	7.59	2.41
	After Set 3	60.19	7	0	21.85	15.93	6.11

Note. TPOT = Teaching Pyramid Observation Tool; CCBOT = Class-Wide Challenging Behavior Observation Tool.

^aSome intervals had both low- and high-intensity problem behavior. ^bThere was one missing data session for Bianca in the final data collection point; therefore, only one 45-min session is presented.

level (i.e., percentage of checklist items present) as when her coach was present and collecting data. This can be seen in Figures 1 to 3 by examining the open triangles.

Generalization

Generalization across activities is indicated in Figures 1 to 3 by the closed diamonds. There were mixed results across teachers. Bianca demonstrated strong results for generalization across activities. After she demonstrated acquisition of a set of practices, her generalization data indicate that she used those practices during activities in which she received no coaching at levels above what was observed during baseline. She demonstrated the practices around criterion levels (80% of checklist items) during the generalization activities. Kendra (see Figure 2) demonstrated inconsistent results for generalization across activities. After she acquired a set of practices, she demonstrated the practices in other activities at higher levels than was observed during baseline for two of the skills (set of targeted practices 1 and 3). However, these levels were significantly lower than those observed in the intervention activities. As can be seen in Figure 3, Susan used the practices during the generalization activities above baseline levels but not near criterion levels.

Maintenance

Maintenance data are represented in Figures 1 to 3 by examining the closed circles in each graph after intervention ended on that goal. After coaching on a goal ended,

intermittent data were collected. Kendra maintained all practices at criterion levels, but Bianca and Susan both needed reminders to continue using practices they had already acquired. These reminders are denoted on their graphs by "R" under the abscissa at the time points when these reminders were given. Bianca received one reminder related to her first action plan goal and two reminders related to her second action plan goal. Susan received three reminders related to her first action plan goal and one reminder related to her second action plan goal. These reminders were simply brief statements to the teacher to remind her to continue using practices; coaching was not provided. With these reminders, both Bianca and Susan maintained their use of practices that were targeted for coaching even when coaching ended. Because of the end of the school year, no maintenance data were collected on third set of targeted practices for any of the teachers.

Classroom-Wide Challenging Behavior

The fourth research question addressed the level of classroom-wide challenging behavior when *Pyramid Model* practices were implemented with fidelity. The results of the CCBOT at each point in data collection are presented in Table 1. For both Bianca and Kendra, there was a decrease in classroom-wide challenging behavior from baseline to intervention. In Susan's classroom, levels of challenging behavior were quite variable, with the lowest level observed during baseline and the highest level observed at the end of intervention.

Overall Implementation of the Pyramid Model

The fifth research question addressed whether overall implementation of the *Pyramid Model* (i.e., TPOT scores) increased after teachers received training and coaching on specific *Pyramid Model* practices. TPOT data are presented in Table 1 by total percentage of indicators on the observation and interview questions, number of environmental items present, and number of red flags present. High scores are desired for indicators and environmental items, and low scores are desired for red flags. There was an increase in the percentage of indicators on the TPOT from baseline to final data collection by approximately 21% for Bianca, 3% for Kendra, and 26% for Susan. In all classrooms, the number of environmental items increased. The number of red flags decreased for Bianca and Susan but stayed the same for Kendra.

Social Validity

The final research question related to social validity and was assessed by examining teachers' perspectives on coaching and the sustainability of the *Pyramid Model* practices. Overall, the teachers found the coaching strategies (e.g., observation, feedback, action planning) useful. Although some teachers found coaches' observations stressful, all agreed that the strategy was useful. Each teacher strongly agreed she had a good working relationship with her coach. Teachers agreed that they would continue to use the *Pyramid Model* practices in the future.

Discussion

The primary purpose of this study was to investigate the effects of professional development intervention that included training and coaching with performance feedback on teachers' use of *Pyramid Model* practices and whether teachers generalized and maintained their use of practices. All teachers in this study increased their use of targeted teaching practices. Thus, there were nine demonstrations/replications across the three teachers, establishing a functional relation. Although it was not possible to establish a functional relation in regard to the teachers' generalization and maintenance of practices (due to the design of the study), there were promising results for both. All teachers maintained their practices with only periodic reminders to use the practices. The results related to generalization to other activities were mixed, with strong generalization demonstrated by Bianca and more limited results for the other two teachers.

The second purpose of this study was to measure the impact of teachers' use of practices on a measure of classroom-wide challenging behavior. These data were mixed across teachers. For Bianca and Kendra, there was a

decrease in challenging behavior from baseline to the end of the intervention. For Susan, there was no discernible pattern in the data on classroom-wide challenging behavior. There was an increase in challenging behavior after baseline, which was due primarily to a new student with significant challenging behavior joining the classroom.

The third purpose was to measure whether teachers' overall TPOT scores increased when they received coaching on specific *Pyramid Model* practices. All teachers' TPOT scores improved after receiving coaching in specific practices. This indicates that scores on a global measure can be positively affected when focusing only on a few key practices.

Limitations

There were five limitations to this study. The first limitation is that each coach collected the primary data in her teacher's classroom. This could have introduced bias in data collection. The IOA data address this concern, and percentages were high across teachers. The coach serving as the primary data collector was also problematic in that the teachers might have only used the practices in the presence of their coach. This was addressed by having the alternate observer checks. These data demonstrated that the teachers used practices when the coach was not present at levels similar to those observed when the coach was present.

The second limitation is that generalization across activities was limited in part because of issues related to when the data were collected. This is because there were few opportunities to observe many of the targeted practices at times other than during the primary data collection (i.e., morning). Because of the numerous routine activities necessary in preschool classrooms (e.g., lunch, nap, bathroom) and the variability in schedules in the afternoon, this was an unavoidable limitation. Nonetheless, this limited the evidence of teachers' generalization of practices.

The third limitation relates to the teachers' maintenance of practices. The school year ended before maintenance data could be collected in the third set of targeted practices. Furthermore, the end of the school year prevented us from observing whether the teachers would have continued to use the first and second sets of targeted practices without the occasional reminders.

The fourth limitation relates to the design of the study. Because the coaching intervention was implemented across practices, the coaches only coached on one set of practices at a time and did not address other practices even when they were linked to the set of targeted practices. For example, when working on schedules and routines, the coaches did not address transitions even though those practices could clearly be related. This might have decreased the likelihood of seeing more significant global changes in teacher

practices. Outside the confines of a study, coaching would not have these restraints.

A final limitation was the inconclusive findings related to the effects of the intervention on children's challenging behavior. There are a number of possible reasons why these findings were not more robust. First, it may be that the specific practices that were targeted (e.g., schedules and routines, problem-solving, emotional literacy) were more likely to affect children's engagement and pro-social behaviors which were not measured in the current study. Second, it is possible that these findings were limited because of the difficulty in measuring incidences of classroom-wide challenging behavior including (a) the influence of a single child's behavior on the classroom-wide measure, (b) the lack of stability in young children's behavior, and (c) the influence of the type and quality of activities occurring in the classroom when the observation is conducted. It is also possible that the "dose" of the teachers' implementation of the practices was not sufficient to affect changes in children's behavior.

Finally, coaching teachers around classroom-wide behavior support practices may not be sufficient for reducing the challenging behavior of individual children. In research studies in which teachers have been effective in implementing interventions that result in reductions of challenging behavior, teachers have been coached on the implementation of behavior intervention practices specifically designed for preventing and responding to an individual child's behavior (e.g., Blair, Fox, & Lentini, 2010; Duda, Dunlap, Fox, Lentini, & Clarke, 2004; Smith, Lewis, & Stormont, 2011). In the current study, teachers were trained to use practices with all children in the classroom, and no data were collected on the use of practices in relationship to individual children with behavior challenges. Future studies might include a coaching component to ensure that practices are being used with sufficient dose, frequency, and in a manner that addresses the challenging behavior of individual children.

Implications for Research and Practice

Implications for research and practice relate to how to ensure generalization and maintenance of practices, the intensity of coaching, and the targeted nature of the coaching. In this study, the results related to generalization were mixed. Teachers did not consistently use targeted practices in untrained activities. This indicates that when coaching teachers, it might be necessary to program for generalization by coaching during different times of the day and in different activities and explicitly coaching teachers on how to use the practices in a variety of activities. The maintenance results indicate that teachers can continue to implement practices in the absence of coaching. Two teachers needed only brief reminders to continue using the practices,

and one did not need any reminders. The use of reminders could be translated into practice through the use of periodic "check-ins" with teachers to ensure their continued use of targeted practices.

Intensive coaching was needed for teachers to implement practices at or above criterion. Coaching was provided to the teacher 2 to 3 times per week in person or via email. Across all three sets of targeted practices, Bianca had 26 live sessions and 10 emails, Kendra had 16 live sessions and 8 emails, and Susan had 21 live sessions and 9 emails. This intensity of coaching should be considered when planning future research around coaching and when making decisions about how to allocate professional development resources within a program. It is possible that this dosage might be reduced if a program-wide model was implemented where teachers received other types of support. It is also possible that teachers working on the same practices could be coached in groups. If an entire program was implementing the *Pyramid Model* at a high level of fidelity, there would be more implementation supports for teachers in the form of administrative commitment, fewer competing demands, peer support, family engagement, and support for practices at all levels of the *Pyramid Model* (Fox & Hemmeter, 2009).

As resources for professional development become more limited, efficient coaching models are needed. Although this study provided evidence that an individualized coaching model was effective, it is important to study how coaching can be provided more efficiently without compromising the integrity and effectiveness of the model. Furthermore, research is needed on how to train and support coaches in their use of effective coaching models.

Acknowledgments

The authors thank the teachers for their participation and hard work in this study. The authors acknowledge the contributions of Patricia Snyder and Lise Fox.

Authors' Note

The opinions expressed are those of the authors, not the funding agency.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study would not have been possible without the support provided by the Institute of Education Sciences, National Center for Special Education Research, Grant R324A070212.

References

- Artman-Meeker, K. M., & Hemmeter, M. L. (2013). Effects of training and feedback on teachers' use of classroom preventive practices. *Topics in Early Childhood Special Education*, 33, 112–123.
- Artman-Meeker, K. M., Hemmeter, M. L., & Snyder, P. (2014). Effects of distance coaching on teachers' use of Pyramid Model practices: A pilot study. *Infants & Young Children*, 27, 325–344.
- Barton, E. E., Pribble, L., & Chen, C. (2014). The use of e-mail to deliver performance-based feedback to early childhood practitioners. *Journal of Early Intervention*, 35, 270–297.
- Blair, K. C., Fox, L., & Lentini, R. (2010). Use of positive behavior support to address the challenging behavior of young children within a community early childhood program. *Topics in Early Childhood Special Education*, 30, 68–79.
- Duda, M. A., Dunlap, G., Fox, L., Lentini, R., & Clarke, S. (2004). An experimental evaluation of positive behavior support in a community preschool program. *Topics in Early Childhood Special Education*, 24, 143–155.
- Fixsen, D. L., Blase, K. A., Duda, M. A., Naoom, S. F., & Van Dyke, M. (2010). Implementation of evidence-based treatments for children and adolescents: Research findings and their implications for the future. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed., pp. 435–450). New York, NY: Guilford Press.
- Fox, L., Dunlap, G., Hemmeter, M. L., Joseph, G. E., & Strain, P. S. (2003). The Teaching Pyramid: A model for supporting social competence and preventing challenging behavior in young children. *Young Children*, 58(4), 48–52.
- Fox, L., & Hemmeter, M. L. (2009). A program-wide model for supporting social emotional development and addressing challenging behavior in early childhood settings. In W. Sailor, G. Dunlap, G. Sugai, & R. Horner (Eds.), *Handbook of positive behavior support* (pp. 177–202). New York, NY: Springer.
- Fox, L., Hemmeter, M. L., & Snyder, P. S. (2014). *Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT™)*, Research Edition. Baltimore, MD: Paul H. Brookes.
- Fox, L., Hemmeter, M. L., Snyder, P. S., Binder, D. P., & Clarke, S. (2011). Coaching early childhood special educators to implement a comprehensive model for promoting young children's social competence. *Topics in Early Childhood Special Education*, 31, 178–192.
- Friedman-Krauss, A. H., Raver, C. C., Neuspiel, J. M., & Kinsel, J. (2014). Child behavior problems, teacher executive functions, and teacher stress in Head Start classrooms. *Early Education and Development*, 25, 681–702.
- Fullerton, E. K., Conroy, M. A., & Correa, V. I. (2009). Early childhood teachers' use of specific praise statements with young children at risk for behavioral disorders. *Behavioral Disorders*, 34, 118–135.
- Gast, D. L., & Ledford, J. (2010). Multiple baseline and multiple probe designs. In D. L. Gast (Ed.), *Single-subject research methodology in behavioral sciences* (pp. 276–328). New York, NY: Routledge.
- Gilliam, W. (2005). *Prekindergartners left behind: Expulsion rates in state prekindergarten systems*. New Haven, CT: Child Study Center, Yale University.
- Hemmeter, M. L., Fox, L., & Snyder, P. (2008a). *Class-Wide Challenging Behavior Observation Tool* (Instrument and manual in preparation).
- Hemmeter, M. L., Fox, L., & Snyder, P. (2008b). *Teaching Pyramid Observation Tool: Pilot version* (Unpublished instrument and manual).
- 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). Baltimore, MD: Paul H. Brookes.
- Hemmeter, M. L., Snyder, P., Kinder, K., & Artman, K. (2011). Impact of performance feedback delivered via electronic mail on preschool teachers' use of descriptive praise. *Early Childhood Research Quarterly*, 26, 96–109.
- Qi, C., & Kaiser, A. (2003). Behavior problems of preschool children from low-income families: A review of the literature. *Topics in Early Childhood Special Education*, 23, 188–216.
- Smith, S. C., 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, 133–143.
- Snell, M. E., Voorhees, M. D., Berlin, R. A., Stanton-Chapman, T. L., Hadden, S., & McCarty, J. (2012). Use of interview and observation to clarify reported practices of Head Start staff concerning problem behavior: Implications for programs and training. *Journal of Positive Behavior Interventions*, 14, 108–117.
- Snyder, P., Hemmeter, M. L., Artman Meeker, K., Kinder, K., Pasia, C., & McLaughlin, T. (2012). Characterizing key features of the early childhood professional development literature. *Infants & Young Children*, 25, 188–212.
- Stormont, M. A., Smith, S. C., & Lewis, T. J. (2007). Teacher implementation of precorrection and praise statements in Head Start classrooms as a component of a program-wide system of positive behavior support. *Journal of Behavioral Education*, 16, 280–290.
- Tapp, J. T., Wehby, J. H., & Ellis, D. (1995). MOOSSES: A multi-option observation system for experimental studies. *Behavior Research Methods, Instruments, & Computers*, 27, 25–31.
- Upshur, C., Wenz-Gross, M., & Reed, G. (2009). A pilot study of early childhood mental health consultation for children with behavioral problems in preschool. *Early Childhood Research Quarterly*, 24, 29–45.