AJSLP

Research Article

Small Groups, Big Gains: Efficacy of a Tier 2 Phonological Awareness Intervention With Preschoolers With Early Literacy Deficits

Lydia G. Kruse,[a] Trina D. Spencer,[b] Arnold Olszewski,[c] and Howard Goldstein[c]

Purpose: The purpose of the present study was to evaluate the efficacy of a phonological awareness (PA) intervention, designed for Tier 2 instruction in a Response to Intervention (RTI) model, delivered to small groups of preschoolers.
Method: A multiple-baseline design across participants was used to evaluate the efficacy of the intervention on low-income preschool children's PA skills. A trained interventionist delivered small group sessions 3 to 4 days a week and ensured children received frequent opportunities to respond and contingent feedback. Participants received 28 to 36 lessons that lasted about 10 min each and focused on PA and alphabet knowledge. Initiation of intervention was staggered across 3 triads, and 7 children completed the study.
Results: The intervention produced consistent gains on weekly progress monitoring assessments of the primary outcome measure for first sound identification (First Sound Fluency). Most children also demonstrated gains on other measures of PA and alphabet knowledge.
Conclusions: Results provide support for the application of a small group intervention consistent with an RTI framework and document the potential benefits of the intervention to learners who need early literacy instruction beyond the core curriculum.

Learning to read may be one of the most important skills that children accomplish. As such, persistent reading deficits observed among school-age children demand attention. For example, nearly two thirds of fourth graders do not read at grade level, and this trend has persisted for years (National Center for Education Statistics, 2011). Fundamental skills necessary for learning to read, such as phonological awareness (PA), develop early in life and are predictive of reading outcomes (e.g., National Early Literacy Panel [NELP], 2008; Storch & Whitehurst, 2002). Weakness in PA skills is associated with difficulty reading (Ehri et al., 2001), and many children, especially those from low socioeconomic status, exhibit deficits in PA (McDowell, Lonigan, & Goldstein, 2007). Given this evidence, interventions that address the development of early literacy skills of young children with identified deficits are critical for promoting long-term literacy skills.

PA has been defined as the "ability to detect and manipulate the sound structure of words independent of their meaning" (Phillips, Clancy-Manchetti, & Lonigan, 2008, p. 3) and is reflected by several skills (e.g., blending, segmenting, rhyming, phoneme isolation), which tend to develop sequentially. In addition, the development of PA skills typically progresses from awareness of larger units (e.g., words in compound words) to syllables to awareness of even smaller units, until children develop phonemic awareness (Lonigan, Burgess, & Anthony, 2000). Phonemic awareness is only one component of PA and describes the specific ability to "focus on and manipulate phonemes in words" (Ehri et al., 2001, p. 253). PA consistently predicts reading outcomes, including the rate at which children acquire reading skills (e.g., Storch & Whitehurst, 2002; Wagner & Torgesen, 1987) with phonemic awareness being one of the best predictors of children's ability to read (Ehri et al., 2001). Despite being fundamental to reading success, PA does not seem to develop naturally (Wagner & Torgesen, 1987). Indeed, PA "is not an intuitive or naturally developing ability … but rather may require deliberate teaching and practice opportunities"

[a]The Ohio State University, Columbus
[b]Northern Arizona University, Flagstaff
[c]University of South Florida, Tampa
Correspondence to Lydia G. Kruse: kruse.101@buckeyemail.osu.edu

Editor: Krista Wilkinson
Associate Editor: Julie Wolter
Received March 13, 2014
Revision received August 11, 2014
Accepted December 30, 2014
DOI: 10.1044/2015_AJSLP-14-0035

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

(Phillips et al. 2008, p. 4). Given mounting evidence of the importance of PA acquisition, federal legislation regarding prevention of reading difficulties (e.g., No Child Left Behind), and widespread use of generalized outcome measures (e.g., Dynamic Indicators of Basic Early Literacy Skills [DIBELS]; Good, Gruba, & Kaminski, 2002), there is a growing expectation that young children master at least some PA skills.

To address the literacy development of all young children, both prevention and intervention efforts are necessary, and a multitier system of support model of service delivery, commonly referred to as Response to Intervention (RTI), is a model that prioritizes both. RTI is a "comprehensive early detection and prevention strategy that identifies struggling students and assists them before they fall behind" (Gersten et al., 2008, p. 4). RTI typically is represented graphically as a pyramid divided into tiers (e.g., three; Fletcher & Vaughn, 2009). These tiers signify levels of increasing instruction (moving up the pyramid) and the approximate percentage of students associated with each level. The lowest level of the pyramid (Tier 1) represents the general education curriculum that applies to all children. The peak of the pyramid (Tier 3) represents the most intense, individualized services, which are reserved for those few students for whom other services did not result in adequate progress. Between Tiers 1 and 3 is secondary tier instruction, intended for students who need supplemental instruction beyond the general education curriculum (Gersten et al., 2008). That is, children whose progress in Tier 1 does not meet an expected level on the basis of screening and progress-monitoring data should receive daily, small group instruction (i.e., Tier 2; Fletcher & Vaughn, 2009). Only a limited number of children in a given classroom are likely to require Tier 2 support if Tier 1 instruction is sufficient (Buysse & Peisner-Feinberg, 2010).

The application of RTI to preschool classrooms is both logical, given the focus on prevention and early intervention, and practical, given the focus on progress monitoring and quality learning environments (Greenwood et al., 2011). Recent survey data indicate a growing trend toward RTI implementation in early childhood learning settings (Greenwood et al., 2011). This trend has sparked interest from practitioners in how to identify children in need of additional instruction and in how to implement tiered instruction, in particular, with efficacious Tier 2 curricula that teach early literacy skills. Children who require Tier 2 interventions would benefit from strategic instruction in small groups on a regular, ongoing basis during which they have frequent opportunities to practice skills associated with later literacy outcomes, such as word segmentation and initial sound identification (Greenwood et al., 2011).

Recent advances in early literacy screening and generalized outcome measurement (e.g., Individual Growth and Development Indicators [IGDIs]: McConnell, Priest, Davis, & McEvoy, 2002; DIBELS: Dynamic Measurement Group, 2006) provide educators with information about students' skills that are known predictors of reading success. Equipped with data regarding children's early literacy skills, educators are better able to determine which students require tiered interventions. For instance, these measures are able to identify students who require additional phonological awareness instruction and the most appropriate level of instruction (e.g., Tier 2).

Although interventions to promote early literacy have been the focus of many ongoing research efforts, their application to Tier 2 instruction in preschool is limited for several reasons. First, many interventions appear to serve as Tier 1 or Tier 3 instruction instead of Tier 2 instruction. For example, many interventions were conducted classwide (e.g., Nancollis, Lawrie, & Dodd, 2005) or delivered to individual children (e.g., Bowyer-Crane et al., 2008; Castiglioni-Spalten & Ehri, 2003) instead of to small groups of children. Second, many intervention programs required lengthy daily training sessions (20–30 min) or many weeks of implementation (e.g., Bowyer-Crane et al., 2008; Castiglioni-Spalten & Ehri, 2003; Justice, Chow, Capellini, Flanigan, & Colton, 2003) that would likely make the interventions difficult to implement in preschool settings. Third, many studies did not specify that participants had PA deficits as an inclusion criterion (e.g., Bowyer-Crane et al., 2008), which raises questions about whether participants may require Tier 2 instruction. Finally, many interventions included children in early elementary school instead of preschool (e.g., Castiglioni-Spalten & Ehri, 2003; Torgesen, Morgan, & Davis, 1992).

Some small group PA intervention studies applicable to preschool Tier 2 instruction have shown some positive effects (e.g., Justice et al., 2003; Koutsoftas, Harmon, & Gray, 2009; O'Connor, Jenkins, Leicester, & Slocum, 1993; van Kleeck, Gillam, & McFadden, 1998). For example, van Kleeck et al. (1998) reported that children in a treatment group that received small group instruction twice a week for 24 weeks demonstrated phonemic awareness skills well above the comparison group at posttest. Likewise, Justice et al. (2003) observed PA skill growth following an emergent literacy intervention in which children participated in twice weekly, 30-min small group intervention sessions for 12 weeks. Children's skills improved significantly on phonological segmentation (and other emergent literacy tasks); however, children's phonological segmentation skills also improved in the comparison group (i.e., adult–child storybook reading).

O'Connor et al. (1993) randomly assigned children to small group treatment (focused on rhyming, blending, or segmenting) for 7 weeks. Each treatment group's scores improved on the target skill when familiar words were presented, but there was limited generalization to unfamiliar words or other PA skills. Finally, Koutsoftas et al. (2009) assigned children to treatment groups that received twelve 20- to 25-min intervention sessions (focusing on first sound identification) 2 days a week. The treatment groups' mean scores on a measure of first sound identification increased after baseline and throughout the intervention; children's phonological segmentation pretest and posttest data confirmed increases in PA skills. The researchers concluded that the intervention was effective; however, some

children "showed small or no treatment effects" (Koutsoftas et al., 2009, p. 122).

These studies testify to the positive effects that PA interventions can have on preschool children's early literacy skills. These studies indicate that some children benefit from focused, small group instruction. Explicit instruction on identified PA skills included "games" and visual materials to increase child engagement, scripts by the interventionist, and modeling and feedback. These components are the foundation for our PA intervention, and the present study extends the previous research in several ways. First, we chose an experimental design, in comparison to a quasi-experimental design (e.g., van Kleeck et al., 1998). Second, we used an inclusion criterion requiring that children demonstrate PA deficits at the onset of the study to focus on children who might benefit from Tier 2 instruction, which was not always the case in previous studies (e.g., Justice et al., 2003; van Kleeck et al., 1998). Third, we prioritized the use of published measures of PA for ease of interpretability and generalization compared to other interventions evaluated by researcher-developed measures (e.g., O'Connor et al., 1993; van Kleeck et al., 1998). Finally, the PA intervention was designed to be brief and delivered daily, which stands in contrast to interventions that were delivered only twice a week and required longer (e.g., 20–30 min) sessions (e.g., Justice et al., 2003; Koutsoftas, et al., 2009).

The purpose of the present study was to evaluate the efficacy of a PA intervention designed for Tier 2 instruction delivered to small groups of preschoolers. The study intent and design were based on evidence that learning to read requires acquisition of code-focused skills. Namely, children must become sensitive to the sound structure of words, and children must make the connection between sounds and letters (Adams, 1990). In comparison to other PA interventions, the intervention under investigation had three features that had the potential to extend the literature.

First, the intervention was designed for use with preschool children. PA interventions delivered in preschool may result in improved school readiness skills and later reading achievement (NELP, 2008). Second, our intervention was designed to be delivered in small groups for children who would benefit from targeted instruction. To date, many PA interventions were designed to be used with individual students or with large groups or require lengthy sessions by a trained professional. Given the limited evidence of the efficacy of PA instruction conducted in large groups (Lonigan, Allan, & Lerner, 2011) and because individualized instruction (especially by a trained professional) may be unrealistic in early childhood settings, our small group (Tier 2) intervention designed for teacher implementation appears to be a logical solution.

Third, the instructional sequence and teaching strategies are grounded in research. For example, the intervention curriculum was sequenced so that larger components of words (e.g., syllables) are taught first, followed by increasingly smaller components (e.g., phoneme), which is consistent with how children acquire PA (Lonigan et al., 2000). Likewise, these skills are introduced in a sequence from (1) blending to (2) segmenting to (3) first part/sound identification that is consistent with evidence on the typical development of PA (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003). These skills are critical for literacy development and associated with greater preventive effects and reading achievement than rhyming skills (Gillon, 2005; Muter, Hulme, Snowling, & Taylor, 1997; Nancollis et al., 2005).

Researchers also suggest that PA instruction be coupled with alphabet knowledge instruction (Ehri et al., 2001; Justice et al., 2003) because neither alphabet knowledge nor PA learned in isolation is sufficient for learning to read, and teaching alphabet knowledge in combination with PA skills results in maximum literacy outcomes for children (NELP, 2008). Therefore, each unit of the PA intervention introduces a new letter and its sound, and all lessons review previously taught letters and sounds. Lessons later in the sequence (e.g., Lesson 7) include activities with printed words and instruction that connects sounds of spoken words to printed letters.

In sum, our relatively brief, 24- to 36-day intervention was designed to teach three important PA skills (i.e., segmenting, blending, and first part/sound identification) and alphabet knowledge. We hypothesized that the PA intervention delivered by an adult in small groups would improve the PA skills of preschool children with identified early literacy deficits. The goals of the study are twofold. The primary goal of the present study was to evaluate the effects of the PA intervention on child outcomes, specifically:

> *Research Question 1*: To what extent does the Tier 2 PA intervention improve the PA skills of preschoolers with identified early literacy deficits, assessed through proximal measures of first parts and sounds of words?
> *Research Question 2*: Do pretest to posttest gains on distal measures of PA support primary findings and indicate generalization of PA?

The secondary goal was to evaluate the feasibility and acceptability of the intervention on the basis of teacher feedback, specifically:

> *Research Question 3*: Do preschool teachers rate the intervention as feasible to implement and having high utility in their classrooms?

We anticipated a delayed effect on progress monitoring measures due to the alignment of the measures with the instructional sequence of the intervention. Gains were expected on distal outcome measures that aligned with our instruction (i.e., Sound ID IGDI) but not necessarily on measures of other PA skills (i.e., Rhyming IGDI, Test of Preschool Early Literacy [TOPEL]; Lonigan, Wagner, Torgesen, & Rashotte, 2007). We expected feasibility and acceptability data to provide important preliminary data in establishing the utility of an evidence-based intervention and its potential for subsequent scale up (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005; Robey, 2004).

Method

Participants

Children attending three Head Start preschool classrooms in an urban setting in the Midwest served as participants. To attend Head Start, all children's families met income eligibility criteria. Each classroom had one lead teacher and at least one assistant teacher. All three lead teachers had earned a four-year college degree and indicated that they used a published curriculum (e.g., The Creative Curriculum), which recommended teaching of early literacy skills, including PA skills. Children in two classrooms (Triads A and C) attended preschool 5 days a week for 6 hr each day; children in the third classroom attended preschool 4 days a week for 3.5 hr each day (Triad B). Each classroom's daily attendance averaged between 15 and 17 children.

Children met the following inclusion criteria: (a) parent consent; (b) regular Head Start attendance (as observed during screening and as reported by classroom teachers); (c) vision and hearing within typical limits (according to Head Start screenings); (d) at least 4 years of age at the start of the study; and (d) demonstration of deficits on PA screening measures (i.e., a score of 5 or less on First Sound Fluency [FSF; Dynamic Measurement Group, 2006], a score of 10 or less on the Rhyming IGDI, and a pattern of deficits during baseline phase). These screening measures provided evidence that the children required Tier 2 PA intervention for several reasons. First, although rhyming was not a targeted skill of the PA intervention, we used the Rhyming IGDI for screening because it was developed to identify children who might require additional instruction beyond the general curriculum (Bradfield, McConnell, Rodriguez, & Wackerle-Hollman, 2013) and as a predictor of later developing PA skills. Second, we excluded some children from the study because their scores on the Rhyming IGDI and FSF were too high, which suggests that we excluded children who were benefiting from Tier 1 instruction. Third, our inclusion criteria were based on converging evidence of (a) low scores on two measures (Rhyming IGDI and FSF) and (b) low FSF scores over a period of time of exposure to Tier 1 instruction (i.e., screening through baseline; 2–3 months).

Nine preschool children (7 girls, 2 boys) attending three different preschool classrooms qualified for inclusion in the study. Participants' demographic and background information was obtained through a survey provided to parents/guardians with the study permission form. Participant characteristics and screening scores are summarized in Table 1. The children's ages ranged from 48 months to 59 months with a mean age of 51 months. Six parents reported the ethnicity of their children as African American, two as White, and one as Latino. Two parents indicated that a second language other than English was spoken in the home, although these children's teachers reported that they spoke English proficiently in the classroom. All but one of the parents reported having completed high school, and two parents had some postsecondary education. No parent reported any concerns with his or her child's development, and according to the children's teachers, no participants received special education services. One participant (Rose) moved during the course of the study, and one participant (Claire) was removed from the study after repeated refusals to participate; therefore, seven children completed the entire study.

Setting

All testing and intervention sessions occurred in a room or hallway near the children's classrooms. Testing was conducted individually, and interventions were delivered in small groups of three children. Thus, a triad was enrolled in each of three Head Start classrooms (in the case of Triads A and B, only two of the three children in each group completed the study; however, we refer to the grouping of children as a triad to reflect the experimental design). At times, there were distractions (e.g., children walking in the hall) during lessons and testing sessions; however, when this happened, the interventionist or examiner quickly redirected the student(s) to the task. Children participated in the intervention and testing sessions during nonliteracy instruction time so that they would not miss important classroom instruction.

Initial screening sessions were conducted at the beginning of the school year. Follow-up screening and the onset of baseline testing occurred approximately 1 month later. Intervention sessions started for the first triad in early November and in early December for the third triad.

Measures

Information about teacher and classroom characteristics was collected using a survey completed by teachers at the end of the study. The survey solicited information about the number of children in the classroom, the number of children with individualized education programs, the number of other adults present in the classroom, length and frequency of preschool sessions, quality and quantity of instruction in the classroom (e.g., curricula used, minutes per day of early literacy instruction), and teacher education and experience. The Clinical Evaluation of Language Fundamentals, Preschool, Second Edition (CELF-2; Wiig, Secord, & Semel, 2004) was administered as a descriptive measure; the TOPEL, Rhyming IGDI, and Sound ID IGDI as screening, pretest, and posttest measures; and FSF and Word Parts Fluency (WPF; Kaminski & Powell-Smith, 2011) as screening and progress monitoring measures.

Descriptive Measure

CELF-2. Participants' general language skills were measured using the CELF-2 (Wiig et al., 2004), which is a standardized, norm-referenced measure of children's language skills. This measure was developed for use with children 3;0 (years;months) to 6;11. Standard scores in the average range fall between 85 and 115. For the purposes of this study, three core subtests were administered (Sentence

Table 1. Participants' demographic characteristics and CELF-2 (screening) scores.

Participant	Gender	Age	Ethnicity	Parent Education	CELF-2 SS
Anne	Girl	53	African American	College degree	86
Eve	Girl	49	African American	HS graduate	75
Jade	Girl	49	African American	HS graduate	79
Kim	Girl	59	African American	HS graduate	81
Liz	Girl	48	African American	Some education after HS	104
Max	Boy	56	African American	HS graduate	88
Sean	Boy	55	Latino	Some HS	65

Note. Age is reported in months and the child's age at the beginning of the study; CELF-2 SS = Clinical Evaluation of Language Fundamentals, Preschool, Second Edition (Wiig et al., 2004) standard score; HS = high school.

Structure, Word Structure, and Expressive Vocabulary) to obtain a core language score. For this measure, the range for internal consistency was reported as .73 to .96 and test–retest reliability for subtests was reported as .77 to .92 (Wiig et al., 2004).

Screening, Pretest, and Posttest Measures

Test of Preschool Early Literacy (TOPEL). Participants' PA and print knowledge skills were evaluated using two subtests of the TOPEL (Lonigan et al., 2007; $\alpha = .93$ [Print Knowledge] and $\alpha = .86$ [Phonological Awareness]). The TOPEL is a standardized measure of print knowledge, vocabulary, and PA ($M = 100$; $SD = 15$). Only the PA and Print Knowledge subtests were administered. The Print Knowledge subtest consists of items related to letter knowledge, letter–sound correspondence, and the use of print in text. The PA subtest consists of items related to blending and elision. TOPEL alpha reliability coefficients ranged from .87 to .96; criterion validity estimates ranged from .59 to .77.

Rhyming individual growth and development indicators (IGDI). Participants' rhyme identification skills were assessed using the Rhyming IGDI 2.0 (CEED@UROC, 2011b). The Rhyming IGDI is a 15-item measure that involves the examiner pointing to and naming three or four pictures on the card, then asking the child to identify which words (or pictures) rhymed (i.e., "Bees, cheese, cat. Which two rhyme?"). This measure is untimed and has a maximum score of 15. The reported estimate of internal consistency on the basis of congeneric reliability was 0.90 (Bradfield et al., 2013). Concurrent construct-related validity correlation with the TOPEL PA subtest is .49.

Sound ID IGDI. Participants' letter–sound correspondence was assessed using the Sound ID IGDI 2.0 (CEED@UROC, 2011a). The Sound ID IGDI is a 15-item measure that involves an examiner showing the child a stimulus card with three letters printed in a row and asking the child which letter makes a target sound (i.e., "Which letter makes the sound /f/?"). This measure was untimed and had a maximum score of 15. The reported estimate of internal consistency on the basis of congeneric reliability was 0.81 (Bradfield et al., 2013). Concurrent construct-related validity correlation with the TOPEL-PA was .71.

Screening and Progress Monitoring Measures

First Sound Fluency (FSF). Participants' first sound fluency skills were measured using a modified version of FSF (Dynamic Measurement Group, 2006). This 1-min task designed originally for kindergartners asks children to produce the first sounds of orally presented, single-syllable words. There are multiple, equivalent probes of the measure. FSF has been reported to have adequate reliability and validity for use with preschoolers (Cummings, Kaminski, Good, & O'Neil, 2011). Children earn 2 points for providing the initial phoneme of a word (e.g., /k/ for cat) and 1 point for the initial blend of initial phonemes of a word (/kæ/); the number of points accumulated in 1 min equals the child's total score (maximum score of 60). Three modifications were made to the FSF measure. We simplified instruction and eliminated feedback during the sample items to reduce the possibility of children learning and/or becoming fatigued given the requirement of repeated testing (up to 20 test sessions per child) in our study design. Second, midway through the study (i.e., Week 8 of treatment for Triad A), we added two sets of sample items to the rotation before test sessions to help to cue the children to the task because the original sample items had been repeated so many times (i.e., 9 times). Finally, late in the study we included three "practice" items using instructional language from the lessons for two participants (Anne and Max) after the examiner modeled the three sample items (as before). These two children seemed to have difficulty transferring learning being demonstrated during instruction to testing sessions. These practice items were intended to help them respond to the assessment stimuli and not to extraneous variables or continue with their pattern of restricted responding. The children did not receive contingent feedback on these practice items or subsequent test items.

Word Parts Fluency (WPF). Participants' initial word parts fluency was assessed using a modified version of WPF (under development at Dynamic Measurement Group; Kaminski & Powell-Smith, 2011). This 1-min task asks children to produce the first parts of orally presented, two-syllable words. Unlike FSF, on WPF children earn 1 point for producing either the first syllable, blend, or phoneme of each word (e.g., /peng/, /pe/, or /p/ for penguin), and the number of points accumulated in 1 min equals the child's total score. The maximum score for this measure is 18. No

benchmarks have been established for this measure yet. WPF and FSF have similar testing for mats and timing requirements; therefore, we made the parallel modifications to WPF as we did to FSF.

Classroom feasibility. Consumer satisfaction and social validity data were obtained from all three teachers at the end of the study using in-person interviews and a 5-question survey. Each question was presented using a 6-point Likert scale (1 = *strongly disagree*, 6 = *strongly agree*); questions included the following: "The children benefited from participating in the intervention," "I observed improvements in the children's early literacy skills during classroom activities," "I observed improvements in the children's early literacy skills during classroom assessments," "The children seemed to enjoy the intervention activities," and "The length of the intervention was appropriate for use with the children." In-person interviews included questions about classroom instruction and potential for Tier 2 interventions in their classrooms.

Measurement protocol and reliability. Two doctoral students in human development administered and scored all measures. Screening, pretesting, and posttest measures were administered in multiple sessions, as needed. All assessment sessions of key measures (i.e., FSF, WPF) were audio-recorded. To calculate scoring reliability for FSF and WPF, 20% of all assessments (from Baseline, Treatment, and Maintenance phases) were randomly selected and scored by an additional trained scorer. Interobserver agreement was calculated on an item level by taking the total number of agreements divided by the total number of agreements plus disagreements, multiplied by 100. Mean interobserver agreement was 95% for FSF and 98% for WPF.

Materials and Instruction

We developed a series of PA lessons teaching blending, segmenting, first part identification, and first sound identification of words (see Appendix for a sample lesson). The intervention consists of 12 units of lessons, with 3 lessons in each unit (e.g., 1a, 1b, and 1c) for a total of 36 lessons (see Table 2). Each unit focuses on a new skill, with later units building on skills taught in earlier units. The lessons within a unit contain the same instruction but different instructional items so that children would be exposed to multiple exemplars to promote generalization of the skills. The lessons are designed to be brief (i.e., less than 15 min) and engaging. The instruction across units includes examples of different kinds of words (i.e., compound, one- or two-syllable words, and words with simple and complex initial sounds/parts). Intervention materials include (a) a script for the interventionist to teach the skills, (b) scripted instructions for providing feedback to a group during activities that ask children to respond, and (c) visuals for lesson activities.

All lessons include short games (e.g., "bingo" cards, hand and body movements) to help maintain the children's attention and engagement. Consistent with teaching strategies associated with positive outcomes for children, instruction was planned strategically using models, leads, and tests to prevent errors (Archer & Hughes, 2011). Following each opportunity to respond, the interventionist read scripted feedback contingent on the children's responses. The differential types of feedback included the following: positive (repeating the correct answer if all children provided correct responses), repetition (providing the stimulus again if a child did not respond), or corrective (providing the correct answer and repetition of the stimulus if any child provided an incorrect response). The feedback was delivered to the triad, and for this reason, the most intensive (e.g., repetition or corrective) feedback was always used even if one child responded correctly. In addition, children had frequent opportunities to respond and practice the skill(s) (Archer & Hughes, 2011); children were encouraged to respond frequently both spontaneously (14–40 times) and imitatively (approximately 20 times) during each lesson. Finally, children had opportunities to practice transferring skills at the end of lessons when each child was asked to respond independently to two or three novel, but lesson-relevant, items.

Experimental Design and Conditions

A concurrent multiple-baseline design across participants (Kennedy, 2005) was used to evaluate the effects of the Tier 2 PA intervention on children's first sound identification. The dependent variables were FSF and WPF. Our second research question examining generalization to more distal measures was addressed using a pretest–posttest, within-participants design.

To deliver the intervention in small groups and ensure a valid examination of treatment effects, participants were divided into three triads who moved through baseline, treatment, and maintenance conditions together. We chose this design for two reasons. First, single-case research "allow[s] confirmation of a functional relationship between manipulation of the independent variable and change in the dependent variable" (Horner et al., 2005, p. 168). If gains are observed only after treatment is introduced and similar improvements are replicated across participants, we can accept this as strong evidence for a functional relation between the treatment and behavior change. Confidence that the intervention, with its staggered initiation, is indeed responsible for behavior change increases with each replication. Second, single-case research allows for careful examination of individual differences. For example, differences in baseline performance, the promptness of response to intervention, and the level of mastery may reflect typical variation in a sample. A multiple-baseline design allows us to analyze the nature of each child's progress by inspecting each participant's level, trend, and variability of responding during the treatment condition compared to baseline performance (Horner, Swaminathan, Sugai, & Smolkowski, 2012). Importantly, repeated measurement, which is a distinguishing characteristic of single-case experimental designs, allows for the detection and reduction of threats to internal validity, such as history (e.g., testing environment), repeated

Table 2. Scope and sequence for the phonological awareness intervention units and lessons.

Unit	Skill(s) taught	Example of instructional language
1	Blending compound words and 2-syllable words	Let's say the parts of the word elbow: el (1) bow. Now you say the word.
2	Blending compound words and 2-syllable words, segmenting compound words	Listen to me say a word: rainbow. *(Put hands together.)* Now listen to me say the parts of the word: rain (1) bow. *(Stretch out one hand at a time.)* Say the word rainbow with me: rainbow. *(Put hands together.)* Now let's say the parts of the word: rain (1) bow.
3	Blending 2-syllable words, segmenting compound words and 2-syllable words	Listen to me say the parts of a word: nap (1) kin. Now you say the word.
4	Segmenting 2-syllable words	Let's say the word marble and clap. The word: marble! *(Clap.)* Now let's say the parts of the word: mar (1) ble.
5	Concept of "first," identification of first part of	Watch my fingers and listen to the parts of the word: side (1) walk. *(Hold up one finger then a second finger for each part.)* Say the parts of the word sidewalk with me and hold up your fingers: side (1) walk. *(Hold up one finger then a second finger.)* Now, you say the first part of the word and hold up one finger. (2)
6	2-syllable words	
7	Concept of "sound," identification of little parts of compound and 2-sylllable words, identification of first sound in 1-syllable words	The word sunflower has two big parts: sun and flower. *(Pull strips apart.)* Words also have little parts. Like the word sun. *(Put flower strip aside.)* The little parts of the word sun are /s/ /un/. *(Pull apart word strip cut into the two parts and when put together there is a complete picture of a sun.)* The word: sun. *(Put word strips together.)* The little parts of the word: /s/ /un/.
8	Identification of first sounds (simple) in 1-syllable segmented words	*(Children hold cards with 4 pictures.)* Listen: /m/. Now you point to the one that starts with /m/. Listen: /m/(1) /ud/. What's the first sound /m/ /ud/?
9	Identification of first sounds (complex) in 1- and 2-syllable segmented words	*(Children hold cards with 4 pictures.)* Listen: /tr/. Now you point to the one that starts with /tr/. Listen: /tr/ (1) /ain/. What's the first sound /tr/ /ain/?
10	Identification of first sounds in 1-syllable whole words	Look at these pictures and words *(show card with 3 pictures and printed words)*: cat, hat, bat. These words sound the same but they have different first sounds. Listen: cat, hat, bat *(emphasize first sound)*. I need you to help me figure out the first sounds.
11		Some words have the same first sound. The words bat, bike, and ball all start with /b/. The first sound you hear in bat, bike, and ball is /b/. What's the first sound you hear in bat? (2) Is it /b/ or /m/?
12		This time, let's see how fast you can tell me your answers. I'm going to say some words. You tell me the first sound you hear in the words. Ready? Sled.

testing, or regression toward the mean (e.g., improvement related to extremely low initial scores; Kratochwill et al., 2010).

Following multiple-baseline design conventions, there were three intervention conditions: baseline, treatment (i.e., PA intervention), and maintenance.

Baseline. Baseline data collection was initiated at the same time for all participants (i.e., concurrent multiple-baseline design). During the baseline condition, children were not exposed to any treatment materials but participated in regular classroom instruction. All participants were administered a series of (multiple-probe) baseline measures to evaluate their first sound fluency and word parts fluency performance. Considering that the intervention was delivered to small groups of children, it was necessary to introduce all participants in a triad into the treatment condition at the same time. Typically, the onset of treatment depends on each participant's stable responding in baseline, but in the current study, stability for all participants was a selection criterion (low FSF scores for 2–3 months encompassing initial screening through baseline). An alternate approach that reduces the same threats to internal validity as baseline stability is random assignment of participants to staggered lengths of baseline phases (Kratochwill et al., 2010). Therefore, the three triads were randomly assigned a priori to treatment starting points (and baseline lengths). Triad A, Triad B, and Triad C participated in 3, 6, and 9 baseline sessions, respectively. Approximately two baseline points were collected per week.

Treatment. Following baseline, the PA intervention was introduced in a staggered fashion to classroom triads. The treatment consisted of the interventionist reading the script with lesson instructions. Each lesson lasted 7–15 min. The lessons were not conducted unless at least two of the three children were present on a particular day so that it was always delivered in a small group instead of individually. Children did not receive "makeup" lessons if they were absent on a day when a lesson occurred. Delivery of lessons was somewhat flexible depending on the needs of the triad. That is, the interventionist always delivered the first two versions of each of the 12 units (e.g., 1a and 1b) to each triad, and depending on the participants' mastery of relevant skills, the interventionist determined whether the triad needed the third lesson in that unit (e.g., 1c). Mastery was monitored by the children's performance on the last three instructional items of the "b" version of a lesson. Unless all children in a single triad responded correctly to two of the last three items, the interventionist delivered the third lesson of the unit. All participants were administered progress monitoring assessments (i.e., WPF and FSF) every

4th day during treatment (after three lessons, regardless of which lessons were delivered). The entire treatment condition lasted between 10 and 15 weeks (i.e., 26–38 treatment days and progress monitoring every 4th day). We repeated several lessons (i.e., 10a, 11a, 12a) at the end of the lesson sequence with Triads A and B because participants' scores were low or unstable.

Maintenance. After a period of no treatment following the treatment condition, we conducted three assessment sessions with each child. These test sessions were conducted on different days and approximately 3 weeks after the final lesson was delivered to the respective triad. Three participants (Anne, Kim, and Max) demonstrated inconsistent and/or low performance (relative to their scores during treatment) during maintenance testing, so we collected a fourth data point for them.

Treatment and assessment integrity. Two doctoral students in human development served as interventionists and examiners. Prior to data collection and program implementation, they were trained and demonstrated mastery of skills for the study by "checking out" with senior researchers trained to administer the measures and lessons. All components of the study protocol were described in a manual. Treatment integrity was monitored throughout the study using video recordings. A portion (20%) of the total number of treatment sessions (20/98) was recorded so that an independent observer could evaluate fidelity of treatment and provide corrective feedback to the interventionists, as needed. Trained research assistants reviewed the videos for key intervention criteria (related to setup and implementation) using an eight-item checklist. The mean percentage of steps completed correctly was 97% (range = 88%–100%). Trained research assistants also evaluated the administration fidelity on 20% of WPF and FSF test sessions. A researcher-developed checklist (related to stimuli, timing, and prompts) was used to monitor steps completed correctly. The mean percentage of administration steps completed correctly was 98% (range = 83%–100%) for FSF and 99% (range = 83%–100%) for WPF.

Results

Proximal Measures

Treatment dosage varied by triad (on the basis of the need for the extra "C" lesson or repeated lessons) and participant (depending on attendance). Triad A received 38 lessons, Triad B received 34 lessons, and Triad C received 26 lessons. Five participants (i.e., Anne, Kim, Max, Sean, Liz) participated in 96%–100% of the lessons, but Eve and Jade were in attendance for only 53% and 62% of the lessons, respectively. Figures 1 and 2 present children's FSF and WPF scores during baseline, treatment, and maintenance conditions (separated by solid lines). Phase changes during the treatment condition (e.g., syllable versus phoneme-level instruction) are indicated by dashed lines. Extra data were collected for Anne, Max, and Kim at the end of the treatment and/or maintenance phase because their performance on WPF or FSF was inconsistent. Overall, the figures indicate that all seven children made meaningful gains in FSF, and the five children who had not mastered WPF during baseline improved from baseline to treatment.

Participants demonstrated low and stable baseline performances for FSF, with the exception of Kim. Because Kim's performance on FSF appeared stable and below 10 at the end of baseline, which was the benchmark that we sought to achieve (discussed below), she remained in the study. Although FSF was a key dependent variable, instruction on identification of first sounds of words was initiated not at the onset of the treatment condition but after six units on segmenting and blending. Thus, the initiation of teaching first sounds is indicated by the dashed vertical lines on Figures 1 and 2 corresponding to Unit 7 (see Table 2). Eve, Sean, Jade, and Liz demonstrated clear increases in FSF scores within 1–3 testing sessions following the start of instruction on identification of first sounds. Kim's results were less impressive, because of her higher baseline performance. Improvements were not detected on FSF and WPF for Anne and Max, but their performance during progress monitoring sessions was not consistent with their performance during the lessons. Consequently, changes to the testing language for progress monitoring measures were introduced, as indicated in Figures 1 and 2. Immediately after we made these minor changes, Anne's and Max's scores improved dramatically. All seven participants performed at or above the benchmark for the beginning of kindergarten (i.e., 10; Dynamic Measurement Group, 2010) on FSF during at least three progress monitoring sessions at the end of the treatment condition.

Baseline scores for our second outcome measure, WPF, were higher and less stable than FSF baseline scores. Thus, improvements were not clearly related to the initiation of treatment in a replicable fashion. However, WPF performance may relate to learning to perform the similar FSF task that targets the initial phonemes rather than the initial part of the word. Jade and Kim, and to a lesser extent Eve, demonstrated mastery of the WPF skill with scores at or near the maximum score (18) during baseline. It is worth noting that Jade and Eve were the two participants who showed immediate gains in FSF performance once initial phoneme-level instruction was introduced.

Liz was the only participant to show immediate improvements in WPF once treatment was started. Her improvements also seemed to begin to show some transfer to FSF performance as well. Although this slight upward trend compromises confidence that phoneme-level instruction was solely responsible for improvements in FSF, it was clear that her mastery level was affected as her FSF scores went from 2 to 20 during the last five units of instruction. Three participants (i.e., Anne, Max, Sean) were unable to perform the WPF task correctly during the baseline condition, but their scores indicated growth during PA instruction at the phoneme level.

Figure 1. Participants' First Sound Fluency scores. PA = phonological awareness.

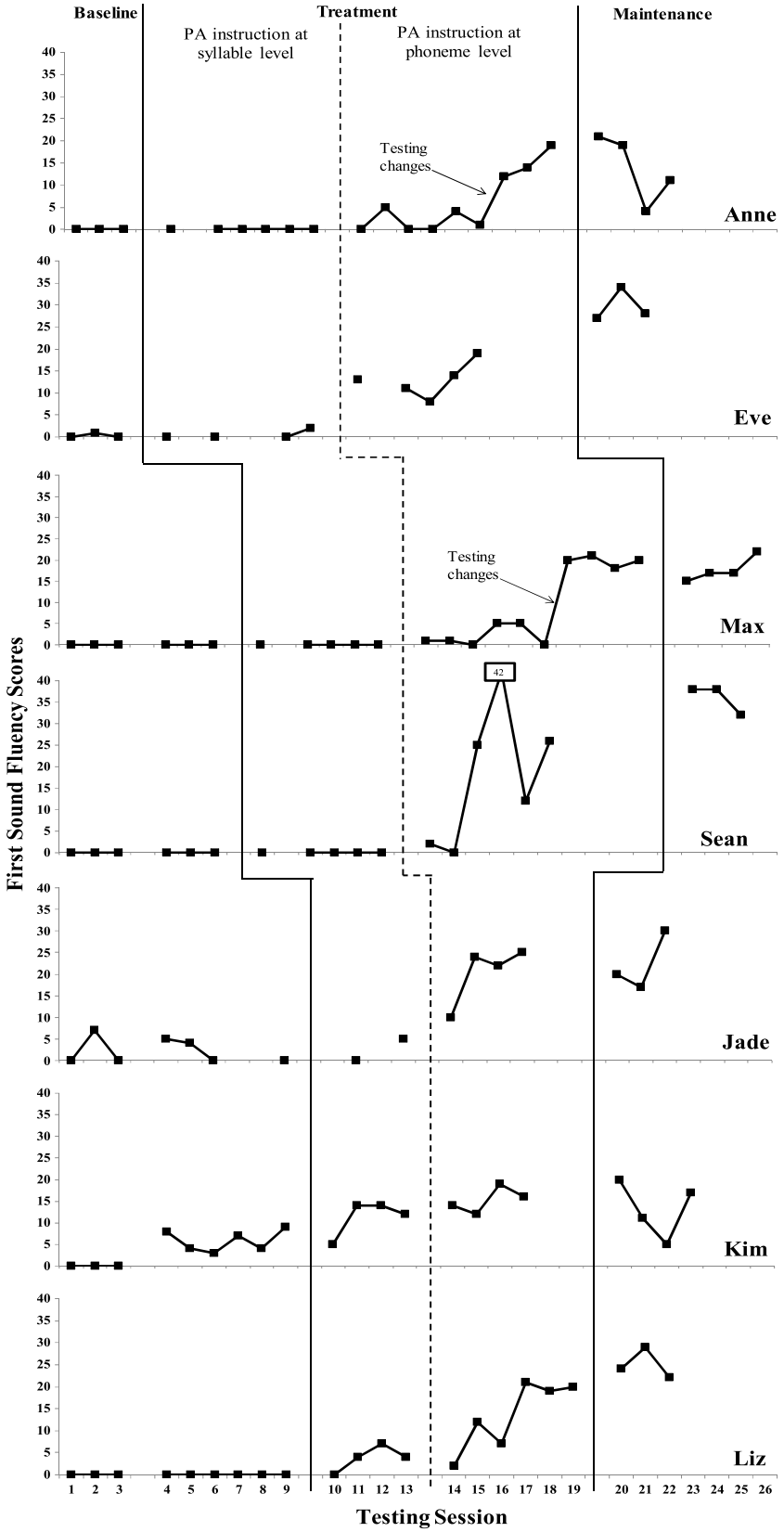
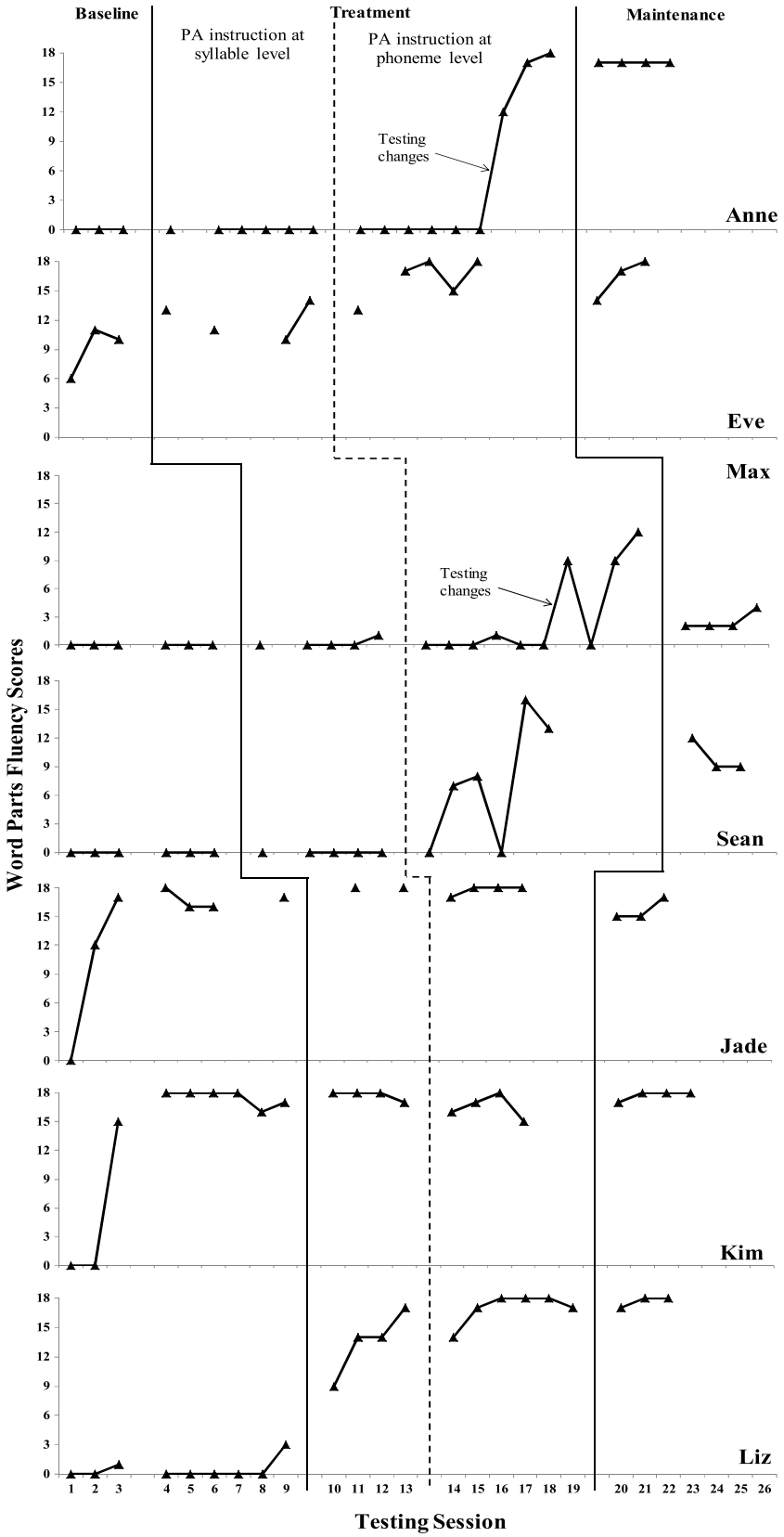


Figure 2. Participants' Word Parts Fluency scores. PA = phonological awareness.



Nonoverlap of all pairs (NAP) was calculated to estimate effect sizes for the multiple-baseline-design data. We chose to use NAP because its application to single-case experimental design has shown advantages over other single-case design overlap-based effect sizes analyses, including high correlation with visual judgment (Parker & Vannest, 2009). NAP was calculated by totaling the overlap of each baseline point (through dotted line for FSF and solid line for WPF) with all treatment and maintenance points divided by total possible overlap pairs. NAP values are classified as "weak" (0–.65), "medium" (.66–.92), or "large/strong" (.93–1.0) according to Parker and Vannest (2009). We observed large effects for FSF for Eve (1.0), Max (.94), Sean (.94), Jade, (1.0), and Liz (.97) and medium effects for Anne and Kim (.88). NAP values for WPF were large for Eve and Liz (.94 and 1.0) and medium for Anne, Max, Sean, and Jade (range = .68–.75). Weak effects were shown for Kim for WPF (.64). The average NAP value for FSF was large (.94) and medium for WPF (.78).

Distal Measures

Six measures were administered at pre- and post-treatment: Rhyming IGDI, Sound ID IGDI, TOPEL (PA and Print Knowledge subtests), WPF, and FSF. Means of baseline scores and treatment scores were calculated separately to evaluate any increase in level. Posttest scores for WPF and FSF were calculated by averaging the final three data points in the treatment condition to improve reliability of the measures. All participants who completed the treatment condition demonstrated discernible gains on WPF and FSF (see Table 3). Six of the seven children made gains of 10 or more on WPF (only Max did not), and all of the children made gains of at least 15 (and as much as 26.7) on FSF. The average gain for WPF was 14.0 and 18.6 for FSF. In general, mean scores confirm gains from pre- to post-treatment for all participants on both measures. Each participant's mean maintenance score for FSF and WPF (see Table 3) was within 3 points of (or higher than) his or her treatment score, with the exception of Max's WPF score. All participants' mean maintenance scores for FSF were above the benchmark for the beginning of kindergarten (i.e., 10).

Six of the seven children who completed the treatment condition improved on the Rhyming IGDI, four improved on the Sound ID IGDI, and four improved on the PA subtest of the TOPEL. Max and Kim's TOPEL PA scores improved 16 and 11 points, respectively; such dramatic gains indicate that these children made more progress than expected with typical development. The average gains were 6.0 for the Rhyming IGDI, 0.4 for the Sound ID IGDI, and 3.1 for the PA subtest of the TOPEL.

Effect sizes were calculated using partial eta-squared, which compared the pretest and posttest means of the within-participants group design. Results indicated significant values for FSF, $F(1, 6) = 107.21$, $p < .01$, $\eta_p^2 = .947$; WPF, $F(1, 6) = 94.695$, $p < .01$, $\eta_p^2 = .932$; the Rhyming IGDI, $F(1, 6) = 9.818$, $p < .05$, $\eta_p^2 = .621$; and the Print Knowledge subtest of the TOPEL, $F(1, 6) = 6.807$, $p < .05$, $\eta_p^2 = .532$. Although these effect sizes may be inflated because of the small n, they would be considered large effect sizes (Fritz, Morris, & Richler, 2012). Effects were not significant for the Sound ID IGDI, $F(1, 6) = .129$, $p = .732$, $\eta_p^2 = .021$; or the PA subtest of the TOPEL, $F(1, 6) = .916$, $p = .375$, $\eta_p^2 = .132$.

Consumer Satisfaction and Social Validity

Consumer satisfaction and social validity data were collected at the end of the study through a teacher survey (see Measures section) and in-person interviews with all three preschool teachers. All teachers indicated that they "agreed" or "strongly agreed" with all of the statements, suggesting positive reactions to the intervention. Teachers' responses during the interviews indicated that they felt that content of the intervention was appropriate for use with their

Table 3. Participants' gains on pretest and posttest measures.

	FSF				WPF				Rhyming IGDI			Sound ID IGDI			TOPEL PA			TOPEL PK		
Participant	Pre	Post[a]	Gain	Mnt	Pre	Post[a]	Gain	Mnt	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain	Pre	Post	Gain
Anne	0	15	15	13.8	0	15.7	15.7	17.0	0	10	10	13	14	1	93	96	3	116	124	8
Eve	1	13.7	12.7	29.7	0	17	17	16.3	4	8	4	9	4	–5	90	90	0	92	98	6
Jade	2	23.7	21.7	22.3	4	18	14	15.7	0	14	14	8	7	–1	90	96	6	107	112	5
Kim	0	15.7	15.7	13.3	0	16.7	16.7	17.8	0	8	8	8	7	–1	71	87	16	92	94	2
Liz	2	20	18	25.0	0	17.7	17.7	17.7	5	7	2	10	12	2	101	93	-8	121	122	1
Max	0	20.7	20.7	17.8	0	7	7	2.5	8	7	–1	5	10	5	74	85	11	77	84	7
Sean	0	26.7	26.7	36.0	0	9.7	9.7	10.0	0	5	5	10	12	2	71	65	–6	80	104	24
M	0.7	19.4	18.6	22.6	0.6	14.5	14.0	13.9	2.4	8.4	6.0	9.0	9.4	0.4	84.3	87.4	3.1	98.8	106.4	8.4
SD	1.0	4.8	4.8	8.4	1.5	4.4	4.1	5.7	3.3	2.9	5.1	2.4	3.6	3.2	12.1	10.8	8.7	15.1	13.9	7.5

Note. Gains are over a period of 28–36 daily treatment sessions. FSF = First Sound Fluency (Dynamic Measurement Group, 2006); WPF = Word Parts Fluency (Kaminski & Powell-Smith, 2011); IGDI = Individual Growth and Development Indicators (CEED@UROC, 2011a, 2011b); TOPEL PA and PK = standard score on Phonemic Awareness and Print Knowledge subtests of the Test of Preschool Early Literacy (Lonigan et al., 2007); Mnt = Mean of maintenance data points.

[a]Mean of final 3 intervention data points.

students and that they would like to implement such an intervention in their classrooms.

Discussion

The purpose of this study was to evaluate the efficacy of a PA intervention delivered by an adult to small groups of preschool children who demonstrated deficits in PA skills. Results supported our primary hypothesis that children identified with deficits in PA would make gains on progress monitoring measures of PA (i.e., FSF and WPF) during the treatment condition. FSF reflected metalinguistic ability at the phonemic level, whereas WPF tended to reflect phonological awareness at the syllable level. Children's gains on FSF were noteworthy. FSF baseline data provided evidence that children's ability to identify first sounds were low and not showing improvement prior to treatment (with the exception of Kim for whom FSF may have been emerging in baseline). As expected, all seven participants demonstrated discernible gains on FSF during the treatment condition. The effects were delayed (as expected) given the instructional sequence (explicit instruction on identification of first sounds was initiated in Unit 7). Data from the second outcome measure, WPF, supported the finding that children's PA skills improved; all four children who had low WPF scores in baseline demonstrated gains during treatment. As expected, children became more fluent on these tasks throughout the treatment condition. Maintenance FSF and WPF data (collected 3–4 weeks after the final lesson) indicated that all children maintained the PA skills that they learned during lessons, with the exception of Max's WPF performance. NAP values indicated medium to strong effects for all children's progress on WPF and FSF with the exception of WPF for Kim. NAP values provide a useful estimate of treatment effect; however, visual inspection of treatment data suggests that some NAP values may be inflated, especially when baselines were highly variable (e.g., WPF for Jade and Kim). In general, the Tier 2 PA intervention contributed to phonemic-level awareness on the basis of improved identification of first sounds among preschool children who showed delays in early literacy development.

Although all children demonstrated posttest gains on the FSF measure, there was some variability in the immediacy and extent of treatment effects. There are several possible explanations for these discrepancies. First, several students (including Kim) exhibited problem behaviors during treatment (e.g., interrupting, getting out of their seat) that may have contributed to inconsistent performance. If treatment were delivered in classrooms by teachers or in the context of a full-scale RTI model, these students would likely receive behavioral supports to improve their focus and reduce the number of distracting incidents during the intervention sessions. Second, children who demonstrated early growth on or mastery of WPF seemed to demonstrate progress on FSF earlier than children whose WPF scores improved midway or later in the treatment condition, an observation that is consistent with the development of PA skills. Other individual variables (e.g., CELF-2 or TOPEL scores, attendance) did not seem predictive of FSF gains. Third, the fact that all children improved on the FSF task to a level exceeding the benchmark for the beginning of kindergarten is all the more impressive as one might not expect all students to respond favorably to a Tier 2 early literacy treatment. For example, Torgesen (2000) examined treatment studies for early literacy and found that between 2% and 6% of students failed to make gains during evidence-based interventions. Such variability among students suggests the need for multiple tiers of instruction in early childhood classrooms. Students who do not respond favorably to Tier 2 treatment may benefit from more comprehensive Tier 1 instruction or more intensive or individualized instruction, such as a Tier 3 intervention.

Results also supported our hypothesis that the children would make gains on some distal measures of early literacy skills. These results should be considered descriptive, as the experimental design did not provide a comparison group for evaluating the meaningfulness of growth on these measures. Nonetheless, partial eta-squared (effect sizes) values indicated large effects on many pre–post measures (e.g., FSF, WPF, TOPEL [Print Knowledge], Rhyming IGDI), seemingly confirming that children made generalized gains in PA. One surprising finding was that children demonstrated gains on the Rhyming IGDI instead of the Sound ID IGDI. We did not anticipate that children's PA skills would generalize to tasks not explicitly taught in the PA intervention. This result could indicate a generalized effect of learning more advanced PA skills (e.g., first sound identification). However, we cannot rule out the possibility that the Rhyming IGDI gains that we observed were a function of typical maturation or classroom instruction. Minimal gains on the Sound ID IGDI could be explained by the fact that the measure tested only a limited number of letter sounds (i.e., 10), and most of these letters did not overlap with the letter sounds taught in the PA intervention. We did not observe substantial gain on the PA subtest of the TOPEL, which may be attributable to a lack of alignment between PA items on the TOPEL and our PA intervention. It appears that the skills gained as a result of the PA intervention did not generalize to PA skills not explicitly taught (e.g., elision). Future research efforts could focus on whether the development of some PA skills (e.g., segmenting) generalizes to other PA skills (e.g., rhyming) in preschoolers. Further research on the development of PA measures to use with preschoolers also is warranted.

Three classroom teachers' feedback indicated that the children in their classrooms benefited from the intervention, including statements that the children "enjoyed participating" and "learned" from the experience. In addition, all teachers "strongly agreed" that the length of the intervention was appropriate. This response boosts our expectation of potential use by teachers in preschool classrooms. We recognize some teachers may have difficulty implementing the intervention in their classes given limited staff, demanding schedules, and minimal experience delivering interventions. However, we are optimistic about classroom

use on the basis of teacher-consumer satisfaction information. Some advantages of the PA intervention include that it is delivered in small groups, lessons are scripted, and each lesson takes less than 15 min to deliver. In addition, optional "C" lessons allow teachers some control in the pacing of instruction, and data can be easily collected at the end of each lesson to regularly inform teachers about children's progress.

Study Strengths

There are several strengths of the study that are worth noting. First, the positive outcomes of our intervention are particularly salient given that all of the participants were from low-income households, a risk factor that has been consistently associated with reading problems (NELP, 2008). Second, the Head Start students selected demonstrated deficits in PA skills. Previous PA intervention studies (e.g., van Kleeck et al., 1998) did not specify participant selection criteria, a factor that calls into question the applicability of their results to a population needing Tier 2 instruction. Our participant-selection criteria suggest a means to identify children who require Tier 2 support, which may inform future early literacy research efforts. Third, our use of a single-case experimental design revealed individual differences in learning trajectories. For example, Sean and Jade made dramatic improvements when instruction on first sounds began, whereas Liz and Kim made more gradual progress. The design also allowed us the flexibility to make modifications during the study; the change in testing procedure for Anne and Max resulted in immediate improvements in their FSF and WPF performance. In general, participants' low and stable FSF baseline performances gave us confidence that our intervention led to the gains we observed, whereas results for WPF were less impressive, because of high scores for three of the children in baseline. Otherwise, the treatment effect was validated through multiple replications across participants and stable maintenance scores.

Two additional strengths of the study are that treatment included instruction at the phoneme level (and not only the syllable level) and that we used FSF as a primary outcome measure. Manipulation of phonemes is strongly associated with later reading achievement (Ehri et al., 2001; Hulme et al., 2002) and is a common indicator of readiness to read (Gillon, 2005). Gains on FSF suggest meaningful PA improvement. Using FSF was somewhat ambitious for our study because it was originally developed for use with kindergarteners. Nevertheless, it was an effective measurement choice for the present study because it has strong psychometric properties, provides benchmark information, is brief, and has a high ceiling score (i.e., 60). FSF data stand in contrast to data from researcher-developed measures reported in many other PA intervention studies (e.g., O'Connor et al., 1993), because using measures with established benchmarks helps consumers to interpret outcomes. All of the participants achieved a FSF performance level expected at the beginning of kindergarten (i.e., 10) on at least three probes, a success that highlights the clinical significance of the intervention.

Limitations and Future Directions

Despite its strengths, our study is not without limitations. One challenge we encountered was related to repeated testing. Despite being a critical component of a single-case experimental design, repeated testing (as many as 21 sessions per child) introduced some adverse effects, including disparate performances during training and progress-monitoring sessions. We observed restricted patterns of responding as some participants fell into a pattern of responding that persisted even when new skills were learned. For example, Anne and Max continued repeating the whole word instead of the first part during progress monitoring even though they were responding during the independent performance portion of lessons. Following some simple modifications to testing sessions, both participants' scores improved, and we concluded that the repeated testing adversely affected their performance in progress-monitoring sessions. Possibly this behavior resulted because (a) there was minimal feedback during assessment sessions, which may have reinforced their pattern of responding, and (b) opportunities for children to respond during lessons were within a context of instruction and models compared to decontextualized prompts during test sessions. A study design with less testing (e.g., randomized control trial or a multiple-probe design) may avoid this problem. Nevertheless, the lack of immediate treatment effects for Anne and Max diminish the confidence that effects were solely due to the intervention.

Another challenge involved the study's outcome measurement. First, two children (Jade and Kim) achieved the ceiling score for WPF in baseline, so we were unable to detect changes with this measure for them during the treatment condition. In the future, more test items could be included for this measure, or the use of alternative scoring might make it more sensitive to phoneme identification. Second, we chose not to use a measure of blending or segmenting because we prioritized using measures with high psychometric properties, and, unfortunately, such progress-monitoring measures of blending and segmenting are not yet available. Consequently, the lack of evidence of children's skill development in blending and segmenting is a limitation of the study because blending and segmenting were early-learning targets in the intervention. Future research should focus on the development and evaluation of different kinds of measures of blending, segmenting, and manipulation of phonemes for use with preschoolers to improve evaluation of interventions and help educators track children's progress. Third, children's TOPEL scores were difficult to interpret. The use of a norm-referenced measure helps us to evaluate the generalized effects of intervention, but unfortunately not all children demonstrated substantial gains on the PA subtest. We caution that the TOPEL scores may not fully reflect children's PA gains because (a) PA activities on the TOPEL focus on blending and elision and not identification of parts and sounds of a word and (b) the TOPEL

focuses on receptive skills compared to FSF, which is a production task. In the future, a different norm-referenced measure might be a better assessment of generalized gains in PA skills.

A final limitation of our study was that trained doctoral students, instead of early childhood educators, implemented the intervention. Although this factor ensured high fidelity of implementation and helped establish efficacy of the PA intervention, it limits generalizability. Despite initial indications of high feasibility and acceptability from teachers (gathered through interviews and a survey), it will be necessary to further evaluate the feasibility when teachers are implementing the intervention in preschool classrooms. Future research also is needed to determine the extent to which success with this intervention prevents reading disabilities, its application to different populations of children, what child factors are predictive of PA outcomes, and whether children maintain the skills beyond the few weeks that we documented.

Clinical Implications

This study has demonstrated the efficacy of a Tier 2 early literacy intervention for preschool students. There are several important implications for speech-language pathologists (SLPs) and other educators. First, results of the study indicate that this PA curriculum was an effective method of promoting PA skills of preschool children with identified early literacy deficits. It shares features with other effective Tier 2 instruction in early literacy in being systematic, explicit, and intensive (Justice, 2006). The scripted nature of this intervention simplifies its implementation for SLPs, early childhood educators, or instructional assistants; the interventionist is provided with simple instructions on how and when to model PA skills. The small group format of delivery provides frequent opportunities for children to practice those modeled skills. The scripted instruction also instructs educators to provide immediate feedback to the group to ensure that appropriate responses are reinforced and that inaccurate responses are corrected. This curriculum also seems viable as daily instruction because lessons are brief and the scope and sequence outlined in Table 2 can be completed in approximately 8 weeks.

Second, this study exemplifies screening procedures that were successful in identifying students who benefited from supplemental instruction in early literacy skill. Also, the brief progress monitoring measures were useful in tracking students' response to intervention. Such assessments are key elements of RTI (Greenwood et al., 2011). In addition, the independent performance probes at the end of each session should help inform interventionists whether this Tier 2 instruction is progressing too slowly or too rapidly for individual children. Within a multitiered system of support, it is possible that progress monitoring every 2–3 weeks may be sufficient to inform a decision-making framework whereby students who are progressing rapidly are reassigned to general instruction and children who are not progressing receive individual instruction.

Finally, SLPs and educators can be pivotal in the implementation and success of RTI in various early childhood settings. For example, by utilizing the curriculum used in this intervention study, SLPs and educators may collaborate to implement a system of screening, intervention, and progress monitoring to ensure that all students in the classroom are mastering fundamental early literacy skills. This Tier 2 intervention represents a relatively efficient means of helping children demonstrating deficits in PA skills to learn phonemic-level early literacy skills that should help prepare them for success in kindergarten and beyond.

Acknowledgments

This work was supported in part by Cooperative Agreement R324C080011, the Center for Response to Intervention in Early Childhood, from the Institute of Education Sciences, U.S. Department of Education.

References

Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 30,* 470–487. doi:10.1598/RRQ.38.4.3

Archer, A. L., & Hughes, C. A. (2011). *Explicit instruction: Effective and efficient teaching*. New York, NY: Guilford Press.

Bowyer-Crane, C., Snowling, M. J., Duff, F. J., Fieldsend, E., Carroll, J. M., Miles, J., . . . Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 49,* 422–432. doi:10.1111/j.1469-7610.2007.01849.x

Bradfield, T., McConnell, S., Rodriguez, M., & Wackerle-Hollman, A. (2013). *Summary of psychometric characteristics for second-generation Individual Growth and Development Indicators for universal screening.* Unpublished technical report, University of Minnesota, Minneapolis.

Castiglioni-Spalten, M. L., & Ehri, L. C. (2003). Phonemic awareness instruction: Contribution of articulatory segmentation to novice beginners' reading and spelling. *Scientific Studies of Reading, 7,* 25–52. doi:10.1207/S1532799XSSR0701_03

CEED@UROC. (2011a). *Fall seasonal identification scale, Individual Growth and Development Indicators for alphabet knowledge: Sound identification.* Unpublished assessment, Center for Response to Intervention in Early Childhood, University of Minnesota, Minneapolis, MN.

CEED@UROC. (2011b). *Fall seasonal identification scale, Individual Growth and Development Indicators for phonological awareness: Rhyming.* Unpublished assessment, Center for Response to Intervention in Early Childhood, University of Minnesota, Minneapolis, MN.

Cummings, K. D., Kaminski, R. A., Good, R., & O'Neil, M. (2011). Assessing phonemic awareness in preschool and kindergarten: Development and initial validation of First Sound Fluency. *Assessment for Effective Intervention, 36,* 94–106. doi:10.1177/1534508410392209

Dynamic Measurement Group. (2006). *First Sound Fluency—Experimental version*. Eugene, OR: Author.

Dynamic Measurement Group. (2010). *DIBELS next benchmark goals and composite scores*. Eugene, OR: Author.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250. doi:10.1598/RRQ.36.3.2

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication No. 231). Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network. Retrieved from http://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf

Fletcher, J., & Vaughn, S. (2009). Response to Intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30–37. doi:10.1111/j.1750-8606.2008.00072.x

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141,* 2–18. doi:10.1037/a0024338

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice guide.* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Gillon, G. (2005). Facilitating phoneme awareness development in 3- and 4-year-old children with speech impairment. *Language, Speech, and Hearing Services in Schools, 36,* 308–324. doi:10.1044/0161-1461(2005/031)

Good, R. H., III, Gruba, J., & Kaminski, R. A. (2002). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (Vol. 1, Vol. 2) (pp. 699–720). Washington, DC: National Association of School Psychologists.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills*. Eugene, OR: Institute for Development of Educational Achievement.

Good, R. H., Kaminski, R. A., Shinn, M., Bratten, J., Shinn, M., Laimon, L., … Flindt, N. (2004). *Technical adequacy and design making utility of DIBELS* (Technical Report No. 7). Eugene: University of Oregon. Retrieved from https://dibels.uoregon.edu/docs/techreports/DIBELS_Technical_Adequacy_TR07.pdf

Greenwood, C. R., Bradfield, T., Kaminski, R., Linas, M., Carta, J. J., & Nylander, D. (2011). The Response to Intervention (RTI) approach in early childhood. *Focus on Exceptional Children, 43*(9), 1–22. Retrieved from http://www.crtiec.org/aboutcrtiec/documents/FOEC-V43-9-May-v5.pdf

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179. doi:10.1177/001440290507100203

Horner, R., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Considerations for the systematic analysis and use of single-case research. *Education and Treatment of Children, 35,* 269–290. doi:10.1353/etc.2012.0011

Hulme, C., Hatcher, P., Nation, K., Brown, A. M., Adams, J. W., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82,* 2–28. doi:10.1006/jecp.2002.2670

Justice, L. M. (2006). Evidence-based practice, Response to Intervention, and the prevention of reading difficulties. *Language, Speech, and Hearing Services in Schools, 37,* 284–297. doi:10.1044/0161-1461(2006/033)

Justice, L. M., Chow, S., Capellini, C., Flanigan, K., & Colton, S. (2003). Emergent literacy intervention for vulnerable preschoolers: Relative effects of two approaches. *American Journal of Speech-Language Pathology, 12,* 320–332. doi:10.1044/1058-0360(2003/078)

Kaminski, R., & Powell-Smith, K.A. (2011). *Word Parts Fluency*. Unpublished assessment, Dynamic Measurement Group, Eugene, OR.

Kennedy, C. H. (2005). *Single case designs for educational research*. Boston, MA: Allyn and Bacon.

Koutsoftas, A., Harmon, M., & Gray, S. (2009). The effect of Tier 2 intervention for phonemic awareness in a Response-to-Intervention model in low-income preschool classrooms. *Language, Speech, and Hearing Services in Schools, 40,* 116–130. doi:10.1044/0161-1461(2008/07-0101)

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Lonigan, C. J., Allan, N. P., & Lerner, M. D. (2011). Assessment of preschool early literacy skills: Linking children's educational needs with empirically supported instructional activities. *Psychology in the Schools, 48,* 488–501. doi:10.1002/pits.20569

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills: Evidence from a latent variable longitudinal study. *Developmental Psychology, 36,* 596–613. doi:10.1037//0012-1649.36.5.596

Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in two to five year old children. *Journal of Educational Psychology, 90,* 294–311. doi:10.1037//0022-0663.90.2.294

Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *Test of Preschool Early Literacy* (TOPEL). Austin, TX: Pro-Ed.

McConnell, S. R., Priest, J. S., Davis, S. D., & McEvoy, M. A. (2002). Best practices in measuring growth and development for preschool children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (Vol. 1, Vol. 2) (pp. 1231–1246). Washington, DC: National Association of School Psychologists.

McDowell, K. D., Lonigan, C. J., & Goldstein, H. (2007). Relations among socioeconomic status, age, and predictors of phonological awareness. *Journal of Speech, Language, and Hearing Research, 50,* 1079–1092. doi:10.1044/1092-4388(2007/075)

Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology, 65,* 370–396. doi:10.1006/jecp.1996.2365

Nancollis, A., Lawrie, B., & Dodd, B. (2005). Phonological awareness intervention and the acquisition of literacy skills in children from deprived social backgrounds. *Language, Speech, and Hearing Services in Schools, 36,* 325–335. doi:10.1044/0161-1461(2005/032)

National Center for Education Statistics. (2011). *The nation's report card: Reading 2011* (NCES 2012–457). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Early Literacy Panel. (2008). *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Institute for Literacy.

O'Connor, R. E., Jenkins, J. R., Leicester, N., & Slocum, T. (1993). Teaching phonological awareness to young children with learning disabilities. *Exceptional Children, 59,* 532–546.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 20,* 357–367. doi:10.1016/j.beth.2008.10.006

Phillips, B. M., Clancy-Manchetti, J., & Lonigan, C. J. (2008). Successful phonological awareness instruction with preschool children. *Topics in Early Childhood Special Education, 28*(1), 3–17. doi:10.1177/0271121407313813

Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37,* 401–411. doi:10.1016/j.jcomdis.2004.04.003

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38,* 934–947. doi:10.1037//0012-1649.38.6.934

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15*(1), 55–64. doi:10.1207/sldrp1501_6

Torgesen, J. K., Morgan, S. T., & Davis, C. (1992). Effects of two types of phonological awareness training on word learning in kindergarten children. *Journal of Educational Psychology, 84,* 364–370. doi:10.1037//0022-0663.84.3.364

Van Kleeck, A., Gillam, R., & McFadden, T. (1998). A study of classroom-based phonological awareness training for preschoolers with speech and/or language disorders. *American Journal of Speech-Language Pathology, 7,* 65–76. doi:10.1044/1058-0360.0703.65

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 1–21. doi:10.1037//0033-2909.101.2.192

Whitehurst, G. J., Zevenbergen, A. A., Crone, D. A., Schultz, M. D., Velting, O. N., & Fischel, J. E. (1999). Outcomes of an emergent literacy intervention from Head Start through second grade. *Journal of Educational Psychology, 91,* 261–272. doi:10.1037//0022-0663.91.2.261

Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals, Preschool* (2nd ed.). San Antonio, TX: Harcourt Assessment.

Appendix

Sample PA Intervention Lesson

PA Live Lesson 2b
Blending compound words, blending 2-syllable words, segmenting compound words; Letter P introduced
Script Feedback Song ***(instructions for interventionist)***

Reminder: Always give feedback twice, as needed (i.e., if any child does not respond or gives an incorrect response after the first round of feedback). Always give the correct response if any child does not respond or gives an incorrect response after the second round of feedback.

(Show page with P and point to P.) Do you see the red letter? That's the letter P. Say P. (1) The letter P says /p/. Say /p/. (1) What letter is this? ***(Point to P.)*** (1) P! What sound does the letter P make? (1) /p/! Let's sing our song. The letter P says /p/. The letter P says /p/. The letter P says /p/. The letter P says /p/.

Look: popcorn. ***(Show page with popcorn, doorknob, and bedroom, and point to the popcorn.)*** Listen to me say the parts of the word popcorn: pop (1) corn. ***(Stretch out a hand for each part.)*** Now listen to me say the word: popcorn! ***(Clap.)*** Say the parts of the word popcorn with me: pop (1) corn. ***(Stretch out a hand for each part.)*** Now let's say the word: popcorn! ***(Clap.)***

+	Yes. Popcorn.
–/NR	Let's try it again. Say the parts of the word popcorn with me: pop (1) corn. ***(Stretch.)*** Now let's say the word: popcorn! ***(Clap.)***

Look at the doorknob. ***(Point to the doorknob.)*** Say the parts of the word doorknob with me: door (1) knob. ***(Stretch.)*** Now, let's say the word: doorknob. ***(Clap.)***

+	Yes. Doorknob.
–/NR	Doorknob. Let's try it again. Say the parts of the word doorknob with me: door (1) knob. ***(Stretch.)*** Now, let's say the word: doorknob. ***(Clap.)***

Look at the bedroom. ***(Point to the bedroom.)*** Let's say the parts of the word bedroom: bed (1) room. ***(Stretch.)*** Now you say the word. (2)

+	Yes! Bedroom.
–	Bedroom. Let's try it again. The parts of the word: bed (1) room. ***(Stretch.)*** The word: bedroom. ***(Clap.)*** Again. The parts of the word: bed (1) room. ***(Stretch.)*** Now you say the word. (2)
NR	Bedroom. Let's try it again. The parts of the word: bed (1) room. ***(Stretch.)*** Now you say the word. (2)