Article

Hammill Institute on Disabilities

Topics in Early Childhood Special Education
2019, Vol. 38(4) 234–247

DOI: 10.1177/0271121418792300
tecse.sagepub.com
SAGE

Systematic Instruction of Early Math Skills for Preschoolers at Risk for Math Delays

Jessica K. Hardy, PhD[1] and Mary Louise Hemmeter, PhD[2]

Abstract

Early math instruction is the subject of increasing attention by early childhood researchers. There is evidence to suggest that early math skills are highly predictive of later academic achievement but that there are disparities in early math skills among young children. The purpose of this study was to examine the effectiveness of a systematic instructional procedure on early math skills with two preschoolers who were at risk for math delays. Target math skills were selected based on the individual needs of each child and included sorting, patterning, and shape manipulation. The instructional procedure was effective for helping two children acquire early math skills, with a functional relation demonstrated for each participant. There were mixed results for generalization and maintenance. Contributions to the literature, limitations, and implications for research and practice are discussed.

Keywords

at risk for developmental delays/disabilities, disability populations, intervention strategies, single-subject designs, experimental studies, research methodologies

Early math skills are the subject of increasing attention in research on young children's learning and development. For example, the National Research Council, in a report on math in early childhood, recommended that "all early childhood programs should provide high-quality mathematics curricula and instruction" (Cross, Woods, & Schweingruber, 2009, p. 345). This focus on math is driven, in part, by research on the correlation between early math skills and later academic achievement (Sarama & Clements, 2009).

There is growing evidence that early math skills are important predictors of later achievement in math (Jordan, Glutting, & Ramineni, 2010; Jordan, Kaplan, Locuniak, & Ramineni, 2007; Jordan, Kaplan, Ramineni, & Locuniak, 2009), as well as in other academic areas such as reading and science (Claessens & Engel, 2013; Duncan et al., 2007). In a meta-analysis of six longitudinal studies, Duncan and colleagues (2007) found that early math skills were more predictive of later school achievement than early reading skills, early attention skills, and early social-emotional skills. Claessens and Engel (2013) found that early math skills were important predictors of later achievement in math, literacy, and science, as well as grade retention (with lower scores in math predicting grade retention).

There also is research demonstrating disparities in math achievement among children based on socioeconomic status (SES) and race (Denton & West, 2002; Markowitz et al., 2006; Morgan, Farkas, Hillemeier, & Maczuga, 2016). In their analysis of the Early Childhood Longitudinal Study data, Morgan and colleagues (2016) found that children from low SES backgrounds were more likely than children from more economically advantaged backgrounds to have persistent mathematics difficulties in third, fifth, and eighth grades. Thus, there is evidence to suggest (a) mathematics is a highly valued outcome for young children, (b) early math skills are related to later math achievement, and (c) there are significant disparities among young children on measures of math achievement. Therefore, it is necessary to consider how best to provide math instruction to those who are at risk for math delays. This is particularly of concern because of research indicating that children of color and children from low SES backgrounds might receive lower quality instruction (Early et al., 2010).

[1]University of Louisville, KY, USA
[2]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Jessica K. Hardy, Department of Special Education, University of Louisville, Porter Education Building, 1905 S. 1st Street, Louisville, KY 40292, USA.
Email: jessica.k.hardy@gmail.com

Early Math Curricula and Instruction

There are early childhood curricula that focus on math, including Building Blocks (Clements & Sarama, 2007) and Pre-K Mathematics Curriculum (Klein & Starkey, 2002). These curricula have been demonstrated to be effective in increasing children's math knowledge. Building Blocks is effective in improving math outcomes for preschool children as measured by the Research-Based Early Math Assessment (REMA; Clements, Sarama, & Liu, 2008). Clements, Sarama, and Wolfe (2011) found statistically significant differences between the intervention and control groups, with an effect size of 0.72. In an earlier study, there were statistically significant differences between the intervention and control groups (effect size = 1.07), as well as between the intervention and the comparison group, which received a different math curriculum (effect size = 0.47) (Clements & Sarama, 2008). The Pre-K Mathematics Curriculum also has been demonstrated to be effective in improving math outcomes. Researchers examined the effects of the curriculum combined with home-based activities and additional school-based math software and found statistically significant differences in scores on a researcher-developed math assessment for children in the intervention group compared with children in the control group, with an effect size of 0.55 (Klein, Starkey, Clements, Sarama, & Iyer, 2008).

Despite the evidence that early childhood math curricula are effective for increasing math knowledge in young children, there is a need for further research on the specific instructional practices teachers use to teach math skills. Curricula are typically focused on content coverage and activity types, with less attention paid to the specific instructional practices that teachers should use. The use of evidence-based instructional practices is important, as it is becoming increasingly apparent that many young children with disabilities and those who are at risk for delays and disabilities will need individualized and intensive instruction in addition to a universal curriculum (Ochsendorf, 2016).

There is some research on the use of specific instructional practices to teach early math skills. These studies included the use of different types of activities and instructional procedures, including using board games to teach ordinality (Ramani & Siegler, 2008, 2011; Ramani, Siegler, & Hitti, 2012; Siegler & Ramani, 2008, 2009), using structured tasks to teach classification and seriation (Ciancio, Rojas, McMahon, & Pasnak, 2001; Kidd et al., 2012; Pasnak, Greene, Ferguson, & Levit, 2006), and using constant time delay to teach counting, numeral and number word identification, and telling time (Daugherty, Grisham-Brown, & Hemmeter, 2001; Holcombe, Wolery, & Werts, 1993).

In a review of these studies, we found the following instructional strategies were used, in differing amounts and combinations: modeling, prompting, directives, providing information about the skill, and discussion. We also analyzed the studies to determine the feedback strategies that were used, which included reinforcement of correct behaviors, reinforcement of attending behaviors, ignoring incorrect responses, correcting incorrect responses, performing the correct response with the child after the child gives an incorrect response, and modeling the correct response and providing an opportunity for the child to demonstrate it.

In the current study, we built on this prior research by designing an instructional procedure based on the evidence on instructional strategies and feedback strategies used to teach math. For example, the intervention included a prompting component, similar to the studies by Daugherty and colleagues (2001) and Holcombe and colleagues (1993). However, other research indicates the importance of modeling (e.g., Pasnak et al., 2006; Ramani & Siegler, 2008, 2011); providing information about how to perform the skill, especially when it is a more complex skill (e.g., Ciancio et al., 2001); and using an error correction procedure (e.g., Ramani & Siegler, 2008). Thus, the intervention in the current study was designed to include each of these components, which, to our knowledge, have not been used in this specific combination in prior research to teach math skills to preschoolers. Most early math skills are complex, chained behaviors, and we sought to design an intervention to address this complexity. We also sought to operationalize the intervention to allow for future replication, and, as needed, iterative refinement of the intervention. This also allowed us to measure intervention dosage, which is an essential component of understanding the effectiveness and efficiency of an intervention.

In addition, in the current study, we selected targets for each child based on their instructional needs. We screened each child to determine the specific math skills that should be targeted for each child. We then selected three target skills for each child (sorting, patterning, and shape manipulation).

The specific research questions addressed in this study were as follows:

> **Research Question 1:** Is a systematic modeling and prompting procedure effective in helping preschoolers acquire, generalize, and maintain discrete early math skills?
> **Research Question 2:** What is the social validity of the intervention and its effects?

Method

Participants

Three children were recommended by their classroom teacher as meeting the following inclusion criteria: (a) 36 to

Table 1. Participant Characteristics.

Characteristic	Jason	She'quan
Age	4 year, 2 month	4 year, 0 month
Sex	Male	Male
Race	Black	Black
Socioeconomic status	Middle	Low
Mullen early learning composite		
Standard score	109	89
Percentile rank	72	22
Descriptive category	Average	Average
TEAM		
Raw score	16	7
T score	−1.74	−17.25

Note. TEAM = Tools for Early Assessment in Math.

72 months of age, (b) attended school regularly (no more than six absences in the previous 30 school days), (c) maintained attention to adult-directed activities, and (d) demonstrated delays in early math skills. To assess the first criterion, the child's birth date was obtained and a chronological age was calculated. To assess the second criterion, attendance records were examined. The third criterion was assessed solely by teacher report. The fourth criterion was assessed by teacher report and an assessment of each child's math skills using a standardized, norm-referenced assessment, Tools for Early Assessment in Math (TEAM; Clements et al., 2011), described below.

Three children began the study, although one participant dropped out of the child care center and, thus, the study, shortly after the study began. Only data on the remaining two children are presented. Both children were typically developing boys who were 4 years of age at the start of the study and were in the same classroom of their child care center. Jason was a quiet, compliant child who engaged appropriately in all classroom activities and followed teacher expectations. His teacher reported that Jason rarely engaged in math behaviors in the classroom. His abilities, as measured by the Mullen Scales of Early Learning (MSEL; Mullen, 1995), were average for all subscales (visual reception, fine motor, receptive language, and expressive language). Jason's early learning composite standard score was average. She'quan was social with peers and adults. He often needed redirection and reminders of classroom expectations. As with Jason, his teacher reported She'quan rarely engaged in math behaviors. His abilities, as measured by the MSEL (Mullen, 1995) were average for visual reception, fine motor, and receptive language. He was assessed as below average for expressive language. She'quan's early learning composite standard score was average. Jason and She'quan scored well below other 4-year-olds on the TEAM, with *T* scores of –1.74 and –17.25, respectively. In a study of 360 children with an average age of 4.25 years, the mean TEAM *T* score was 44.42 (Clements et al., 2008). An overview of characteristics for both participants is provided in Table 1.

Settings

Both children were in the same classroom in a child care center serving children from low-income backgrounds in a large southern city. The teacher used the *Frog Street Pre-K Curriculum* (Schiller, Flor Ada, Campoy, & Mowry, 2010). The teacher also provided a math center and manipulatives center; however, during a structured observation of the classroom, there was minimal evidence of instruction in math skills. All study activities, except for generalization sessions, occurred at a table in a conference room or staff break room near the children's classroom. Generalization sessions occurred in the classroom, seated at a table or on the floor.

Measures

The MSEL (Mullen, 1995) was administered to provide a measure of each child's developmental status. It was administered by a graduate student trained in its use as part of an unrelated randomized control trial. Administration took 70 to 90 min, distributed across several sessions with breaks included to reduce fatigue. The MSEL is a standardized, norm-referenced measure with five scales (gross motor, visual reception, fine motor, expressive language, and receptive language). All scales except for gross motor were administered to the study participants. These four scales are also combined to derives a composite cognitive score. Individual scales provide *T* scores, percentile ranks, and age equivalents; composite scores provide standard scores and percentile ranks. Normative data were collected on a nationally representative sample of children with no known disabilities. Internal consistency score reliability for the five MSEL scales range from .75 to .83 and for the composite .91. Test–retest score reliability on the composite scale was .84 for children ages 25 to 56 months.

The TEAM (Clements et al., 2011) was used to provide an overall estimate of each child's math knowledge. The TEAM is a standardized, norm-referenced assessment of math skills for children from preschool to second grade. It contains two scales: number and shape. The TEAM was developed using a Rasch model (Clements et al., 2008). This examination provided evidence for the content, face, and concurrent validity of the TEAM (also referred to as the Research-Based Early Maths Assessment; Clements et al., 2008).

On the TEAM, for ease of interpretability, raw scores are translated into *T* scores, and grade equivalents are indicated.

Table 2. Materials for Each Skill.

Skill	Probe	Intervention: Manipulatives	Intervention: Art	Generalization
Sorting (Tier 1)	Different color and size foam shapes (orange, purple, and blue squares, circles, and diamonds)	Different color and size foam shapes (green, yellow, and red; ovals, rectangles, and triangles)	Different color and size construction paper shapes, glue, and paper	Different size and shape puzzle pieces
Patterning (Tier 2)	Colored blocks	Unifix® cubes	Markers and paper strips with repeating pictures	Different color plastic stacking caps
Shape manipulation (Tier 3)	Pattern blocks with picture cards	Pattern blocks with picture cards	Colored paper shapes mimicking pattern blocks, glue, and picture cards	Tangrams and construction paper figures

However, guidance is not given for comparing a child's performance with the typical performance of other children in the same age group. Thus, it is difficult to determine if a child's performance on the TEAM is above or below that of same-aged peers. However, in a study in which the TEAM was developed (Clements et al., 2008), 360 preschool children were assessed using the TEAM at two points in time. At the first time point, their average age was 4.25, roughly equivalent to the children in the current study. In the TEAM development study, the children were from 34 low-income and 12 mixed-income classrooms. The average *T* score among these 360 children was 44.42, with a range of 20.25 to 69.08 and a standard deviation of 7.85. These scores serve as a comparison for the children in the current study. On the TEAM, both participants made several incorrect responses in a row early in the assessment, which necessitated ending the assessment. This meant the TEAM did not provide useful information for determining skills to target for instruction.

The Early Math Screening Instrument (EMSI; Hardy, 2014) was used to screen the children's math skills. The EMSI is a criterion-referenced measure of children's specific early math skills, developed for the purposes of this study. To date, the reliability and validity of the EMSI have not been examined. The EMSI was designed to measure early math skills across all relevant domains, and thus provided more detailed information on multiple math domains than the TEAM. Items on the instrument were developed based on the literature on early childhood math development (e.g., Charlesworth & Lind, 2010; Clements & Sarama, 2009) and standards related to early math, including the Head Start Early Learning Outcomes Framework (Office of Head Start, 2015). The instrument was administered to each child by the first author over three sessions. During each session, the first author administered each item by giving the child materials and a task direction (e.g., "Count the bears"). Each task was assessed in three trials (one per session). Only items that the child was unable to demonstrate in all three probe sessions were considered as potential target skills.

Materials

The materials used for probe, intervention, and generalization sessions all included common, developmentally appropriate manipulatives (e.g., Unifix cubes, miniature vehicles). In addition, in intervention sessions, the participants had the option to do an art activity. The materials used for each skill in each session type are presented in Table 2. Although manipulatives were used in all three session types, they differed slightly so that the materials used in intervention, probe, and generalization sessions were not identical. However, there were some materials that remained constant. For example, for the target skill of shape manipulation, pattern blocks were used in both probe and intervention sessions. The picture cards, however, differed for the two session types.

Due to the scarcity of materials available in the classroom relevant to each skill, all materials for the generalization sessions were provided by the researcher and were brought in and out of the classroom. Materials used in maintenance sessions were the same as were used in probe sessions.

Response Definitions and Measurement

Data were collected live during sessions by the first author using trial recording (Ledford, Lane, & Gast, 2018). Sessions were approximately 10 to 20 min in length. There were four session types: probe, intervention, generalization, and maintenance.

In probe, generalization, and maintenance sessions, the child was given a task direction (e.g., "Put them together by shape"). There were three possible child responses: correct, error, and no response. A *correct response* was recorded if the child physically demonstrated or said the correct answer.

A correct response was also scored if a child initially answered incorrectly and then spontaneously produced the correct response within 5 s. An *error* was recorded if the child physically demonstrated or said a wrong answer after the task direction. A *no response* was recorded if the child did not respond or said, "I don't know."

The exact topography of correct and incorrect responses varied based on the skill. For each skill, correct and incorrect answers were operationalized prior to beginning data collection. For example, for the skill of sorting, the correct response was operationalized as putting all the objects into separate piles based on the attribute named (size or shape). For the skill of patterning, the correct response was operationalized as extending the initial pattern (AB or ABB) at least three times in the correct sequence and then stopping when done (i.e., without adding additional items in an incorrect sequence). For the skill of shape manipulation, the correct response was operationalized as putting the shapes on top of the picture, filling the picture fully, with no extra shapes on the picture.

During instructional sessions, there were two types of trials, demonstration and practice. In demonstration trials, the teacher modeled the target skill but the child was not instructed to respond. In practice trials, a task direction (e.g., "Put them together by shape") was provided. There were three possible child responses in the practice trials: unprompted correct, prompted correct, and prompted error. An *unprompted correct* response was recorded if the child physically demonstrated or said the correct answer after the task direction. When the child did not respond to the task direction or responded incorrectly to the task direction, a prompt was provided. The child's response to the prompt was then coded as either a prompted correct or a prompted error. A *prompted correct* was recorded when a prompt was necessary and the child responded correctly to the prompt. A *prompted error* was recorded when a prompt was necessary and the child responded incorrectly or did not respond to the prompt. For each trial, only one child behavior was coded. Either the initial response was recorded if correct or the response to the prompt was recorded if a prompt was necessary. Therefore, an incorrect response to the task direction was not recorded, as it was followed by a prompt for a correct response.

Interobserver Agreement

Interobserver agreement data were independently collected by two project staff members, who had a master's or doctoral degree in education. Data collectors were trained in practice sessions with nonstudy children prior to the beginning of data collection. Prior to collecting data, data collectors had to reach at least 90% agreement with one another across three consecutive observations for each session type. During the study, interobserver agreement data were collected for at least 30% of sessions for each participant, distributed across session type. Percentage agreement was calculated using the point-by-point agreement method. Total agreements were divided by the sum of agreements and disagreements, which was multiplied by 100 to derive a percentage (Repp, Deitz, Boles, Deitz, & Repp, 1976).

Experimental Design

A multiple probe across skills design (conditions), replicated across participants was used (Gast, Lloyd, & Ledford, 2018). In multiple probe designs, baseline data are collected intermittently, with at least three sessions of consecutive baseline probes prior to the introduction of the independent variable. When baseline data are stable, the independent variable is introduced in a staggered fashion across the three skills. In multiple probe designs, experimental control is demonstrated by stable baseline data and the immediate change in the dependent variable after introduction of the independent variable (Gast et al., 2018). Data in the other tiers should remain stable until the independent variable is introduced in each tier (Gast et al., 2018).

Procedures

Personnel. The first author, who was at the time a doctoral student in early childhood special education, implemented all study activities and data collection, except for the MSEL administration and the interobserver agreement and procedural fidelity data collection. These were implemented by trained research assistants.

Target skill selection. Target skills were selected from a list of possible math skills that were assessed using the EMSI (Hardy, 2014) using the procedures described above. Behaviors that the children could not perform in each of three screening sessions were selected as target skills. The skills chosen for both Jason and She'quan were sorting, patterning, and shape manipulation. These skills were included in the EMSI and selected for instruction based on consensus in the field of early math that they are precursors to later skills in algebra (sorting and patterning) and geometry (shape manipulation).

For each target skill identified, two specific behaviors were taught in an attempt to support more generalized learning than might be expected with just one specific behavior. For the skills of sorting, sorting by shape and by size were the two behaviors taught. For the skill of patterning, the two behaviors taught were completing AB patterns and completing ABB patterns. For the skill of shape manipulation, the two behaviors taught were manipulating three shapes to form pictures and manipulating six shapes to form pictures.

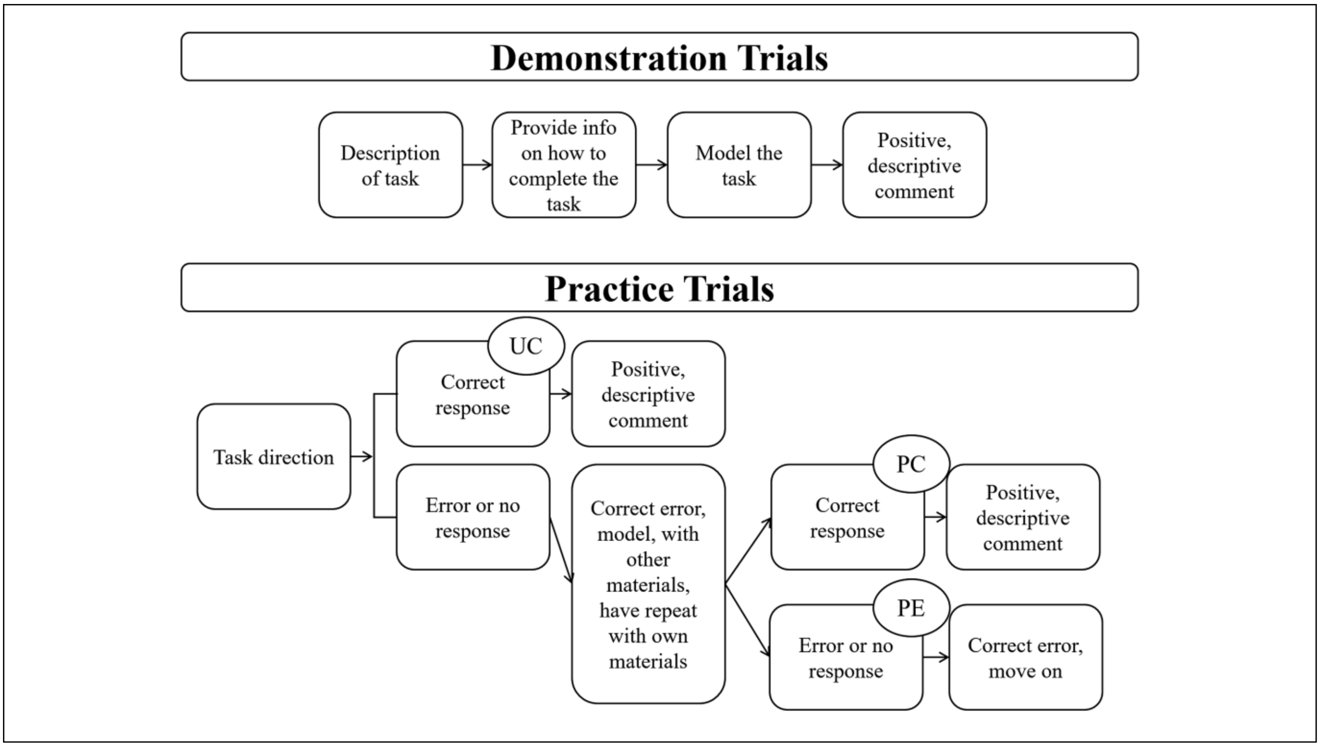

Figure 1. Demonstration and practice trials flowcharts.
Note. UC = unprompted correct; PC = prompted correct; PE = prompted error.

Probe conditions. During the probe condition sessions, the researcher presented the task direction (e.g., "Keep going with the pattern") and then provided a 5 s response interval. The child's response was recorded. The researcher reinforced correct responses and ignored errors and no responses. The child received three trials per target behavior, for a total of six trials per skill and 18 trials per probe session. A probe condition occurred across all three tiers before intervention began in Tier 1. After data were stable and low, intervention in Tier 1 began. Subsequent probe conditions occurred after the child reached criterion in each tier of intervention (three consecutive sessions at 100%). There were four probe conditions over the course of the study.

Intervention. Instruction was provided in the context of working with manipulatives or engaging in art activities (e.g., using markers to color objects in a specific pattern, such as red–green; see Table 2). In each session, the child chose between using the manipulatives or doing an art activity. The art activity option was used for two reasons: (a) to allow the participants to make a choice and thus be more engaged in the session, and (b) to reflect the field's emphasis on play-based, embedded instruction (Division for Early Childhood, 2014). A prompting procedure with two types of trials, demonstration and practice, was used in intervention sessions. In demonstration trials, the researcher provided a description of the task (e.g., "I'm going to put them together by size") and provided information about how to complete the task (e.g., "To put them together by size, you look at each shape to see whether it is little, medium, or big. You put the little ones together, the medium ones together, and then the big ones together"). The researcher then modeled the task, without delivering a task direction. After completing the task, the instructor provided a descriptive comment (e.g., "I put them together by size!"). See Figure 1 for a flowchart of demonstration trials. The child received two demonstration trials for each behavior, for a total of four demonstration trials per instructional session. Each session began with a demonstration trial, and the remaining demonstration trials were interspersed with practice trials so that every third trial was a demonstration trial.

In practice trials, the researcher provided the task direction (e.g., "Put them together by size"). The adult then provided a 5-s response interval. If the child responded correctly, the researcher provided a positive, descriptive comment (e.g., "You did it—you put them together by size!"). This was recorded as an unprompted correct. If the child responded incorrectly, the researcher said, "Almost—let me show you how to do it" (or a similar comment), and covered the child's materials if necessary (e.g., if the child incorrectly extended a pattern using Unifix® cubes, the Unifix® cubes were covered). The researcher then modeled the correct response using different materials. The researcher then uncovered the child's original materials (if they had to be covered) and reconfigured the materials so they were no longer an incorrect example but rather were arranged as they originally were when the trial began. The child was then directed to try again (indicated by the researcher repeating the task direction) using the original materials, with the researcher's model visible. If the child provided a correct response, the researcher provided a positive, descriptive comment, and a prompted correct was recorded. If the child did not produce a correct response, an error was recorded. See Figure 1 for a flowchart of practice trials. The child received eight practice trials per instructional session. Two behaviors were taught, with four trials per behavior. For example, for the skill of sorting, the child received four trials for sorting by shape and four trials for sorting by size. Intervention continued until the child reached criterion, which was three consecutive sessions at 100% unprompted correct.

Intervention modifications. The intervention was modified in two ways for the participants in the course of the study. First, the number of demonstration and practice trials presented to each child was decreased by half because the children were beginning to show signs of fatigue in the intervention sessions. The second modification was a change in the error correction procedure. Originally, the researcher corrected the child's error, modeled the correct response with other materials, and then directed the child to try again with his own materials. This procedure sometimes led to the child being allowed to complete a task incorrectly and then repeat the incorrect response after it was modeled correctly. Thus, the procedure was modified so that the researcher immediately provided the minimal physical and/or verbal assistance needed. For example, if the skill was patterning, and the child incorrectly completed the pattern, the researcher might have said, "The pattern is one white cube, then two green cubes. Put a green cube next." If the child appeared to be trying to correct his error (e.g., said, "Oh wait, let me fix that," or made a movement to undo the error), additional wait time was provided. However, if the child appeared unaware that he made an error or continued making the error, the researcher interrupted the response. This modification was instituted in Tier 2 for each participant.

Generalization. Generalization sessions occurred during probe and intervention conditions approximately every fifth session (once per week). Generalization to the classroom and other materials was measured. The primary researcher conducted the generalization sessions in the child's classroom, using materials that were not used during intervention. The researcher used the same task direction as in practice trials. The child received two probe trials per behavior, for a total of four probe trials for each of the three skills and 12 probe trials per session. The researcher reinforced correct responses and ignored errors and no responses.

Maintenance. Maintenance sessions were conducted approximately 1 month after the end of the last probe condition. Two sessions were conducted. The child received three trials per behavior, for a total of six trials per target skill and 18 trials per session. The researcher reinforced correct responses and ignored errors and no responses. In addition to these maintenance sessions conducted after the completion of all intervention, the probe conditions serve as a measure of maintenance of skills acquired in previous tiers.

Procedural Fidelity

The same data collectors collected interobserver agreement data and procedural fidelity data. A combination of checklists items and trial recording was used to measure the researcher's behavior. Checklists were used to measure researcher behavior that occurred one time per session (e.g., researcher worked with the child in a one-on-one session), and trial recording was used to measure behavior that occurred multiple times in one session (e.g., providing a prompt after the task direction and response interval). Procedural fidelity data were collected in at least 30% of sessions for each participant, distributed across session type. The number of total occurrences was divided by the sum of occurrences and planned occurrences and was multiplied by 100 to derive a percentage (Barton, Meadan-Kaplansky, & Ledford, 2018).

Social Validity

Social validity was measured by asking the teacher to complete a questionnaire about the goals of the study, the procedures used in the study, and the effects of the study on the children's behavior (Wolf, 1978). Because only one teacher was involved in the current study, it was not anonymous. The first set of questions was asked prior to the beginning of data collection and after data collection ended. The teacher was asked if she observed the child engaging in the skills targeted during the intervention. The second set of questions related to whether the teacher viewed early math instruction as important for preschoolers. The third set of questions related to the instructional procedures used in the study. The teacher was asked to view a video of the researcher implementing the intervention and respond to questions about whether the informant would be willing and able to implement the intervention.

Results

Results are presented first for interobserver agreement and procedural fidelity. Then, acquisition, generalization, and maintenance results are presented for each participant, with attention given to changes in level, trend, and variability between conditions. Immediacy of the effect is also considered for each dependent variable for each participant. Next, dosage data are presented across participants. Finally, social validity data are provided.

Interobserver Agreement and Procedural Fidelity

Interobserver agreement and procedural fidelity data were collected in 32.76% of sessions for Jason and 30.38% of sessions for She'quan. The mean interobserver agreement percentages for Jason and She'quan were 99.47% (range = 90%–100%) and 99.13% (range = 87.5%–100%), respectively. The mean procedural fidelity percentages for Jason and She'quan were 99.14% (range = 88.24%–100%) and 99.48% (range = 93.33%–100%), respectively.

Jason

Acquisition of target behaviors. Data on Jason's acquisition of target behaviors are presented in Figure 2. Unprompted correct responses are represented with a closed circle and prompted corrects are represented with an open square. Jason demonstrated none of the target behaviors (sorting, patterning, or shape manipulation) in the probe condition(s) prior to receiving instruction. Jason rapidly acquired sorting behaviors in Tier 1, reaching criterion in six sessions. Jason gradually acquired patterning behaviors in Tier 2, reaching criterion in 23 sessions. His data were variable before reaching criterion. Jason again rapidly acquired shape manipulation behaviors in Tier 3, reaching criterion in six sessions.

Generalization of target behaviors. Data on Jason's generalization of target behaviors to untrained materials in the classroom are presented in Figure 2. Generalization probes are represented with an asterisk. During the first probe condition, Jason demonstrated no sorting behaviors in generalization sessions. By the end of Tier 1, he demonstrated 50% of sorting behaviors in generalization sessions. In Tier 2, he demonstrated an average of 95% of sorting behaviors in generalization sessions, and, in Tier 3, he demonstrated 100% of sorting behaviors in generalization sessions.

Prior to intervention in Tier 2, Jason demonstrated no patterning behaviors when measured in generalization sessions. His use of patterning behaviors in generalization sessions during Tier 2 was variable. In Tier 3, he demonstrated 100% patterning behaviors in generalization sessions.

In the first probe condition, Jason demonstrated no shape manipulation behaviors in generalization sessions. By the end of Tier 1, he began to demonstrate some shape manipulation behaviors in generalization sessions (prior to receiving instruction in these behaviors). He continued to demonstrate these behaviors with 50% accuracy in generalization sessions in Tiers 2 and 3.

Maintenance of target behaviors. Data on Jason's maintenance of target behaviors are presented in Figure 2. Maintenance probes are represented with a closed triangle. Across two maintenance sessions one month after intervention ended, Jason demonstrated 100% of sorting and patterning behaviors and an average of 91.67% of shape manipulation behaviors.

She'quan

Acquisition of target behaviors. Data on She'quan's acquisition of target behaviors are presented in Figure 3. Unprompted correct responses are represented with a closed circle and prompted corrects are represented with an open square. She'quan demonstrated none of the target behaviors (sorting, patterning, or shape manipulation) in the probe condition(s) prior to receiving instruction. She'quan acquired sorting behaviors in Tier 1, reaching criterion in 17 sessions. Throughout the condition, his unprompted corrects had an ascending trend, but there was significant variability. At the beginning of the intervention, his data gradually accelerated to 100% unprompted correct responses in six sessions, and then they were variable before stabilizing and reaching criterion after 17 sessions. In Tier 2, She'quan's patterning data had an ascending trend but never reached mastery criterion. His data also showed significant variability. Although he had three sessions at 100% unprompted correct, they were not consecutive sessions, which was the criterion. The decision was made to move to the next tier when two out of three consecutive sessions were at 100%. She'quan acquired shape manipulation behaviors in Tier 3, reaching criterion in nine sessions.

Generalization of target behaviors. Data on She'quan's generalization of target behaviors to untrained materials in the classroom are presented in Figure 3. Generalization probes are represented with an asterisk. In the first probe condition, She'quan demonstrated no sorting target behaviors in generalization sessions. By the end of Tier 1, he demonstrated 50% of sorting behaviors in generalization sessions. At the beginning of Tier 2, he demonstrated 100% of sorting behaviors in generalization sessions, but by the end of Tier 2, he only demonstrated 50% of sorting behaviors in generalization sessions. In Tier 3, he demonstrated 50% of sorting behaviors in generalization sessions.

Prior to intervention in Tier 2, She'quan demonstrated no patterning behaviors when measured in generalization sessions. His use of patterning behaviors in generalization sessions in Tier 2 was variable, but reached 100% by the end of Tier 2. She'quan's demonstration of patterning behaviors in

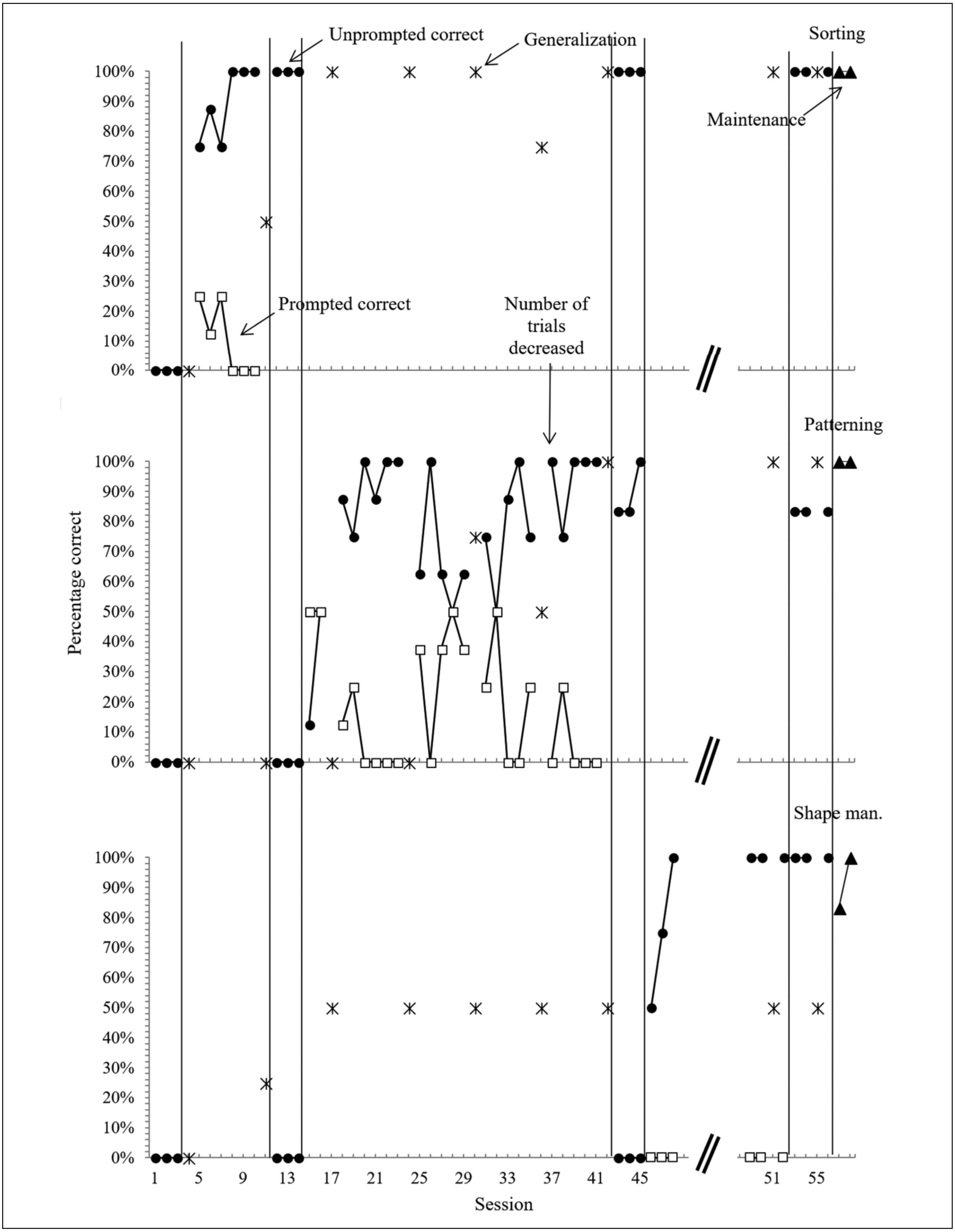

Figure 2. Jason's acquisition, generalization, and maintenance of target behaviors.

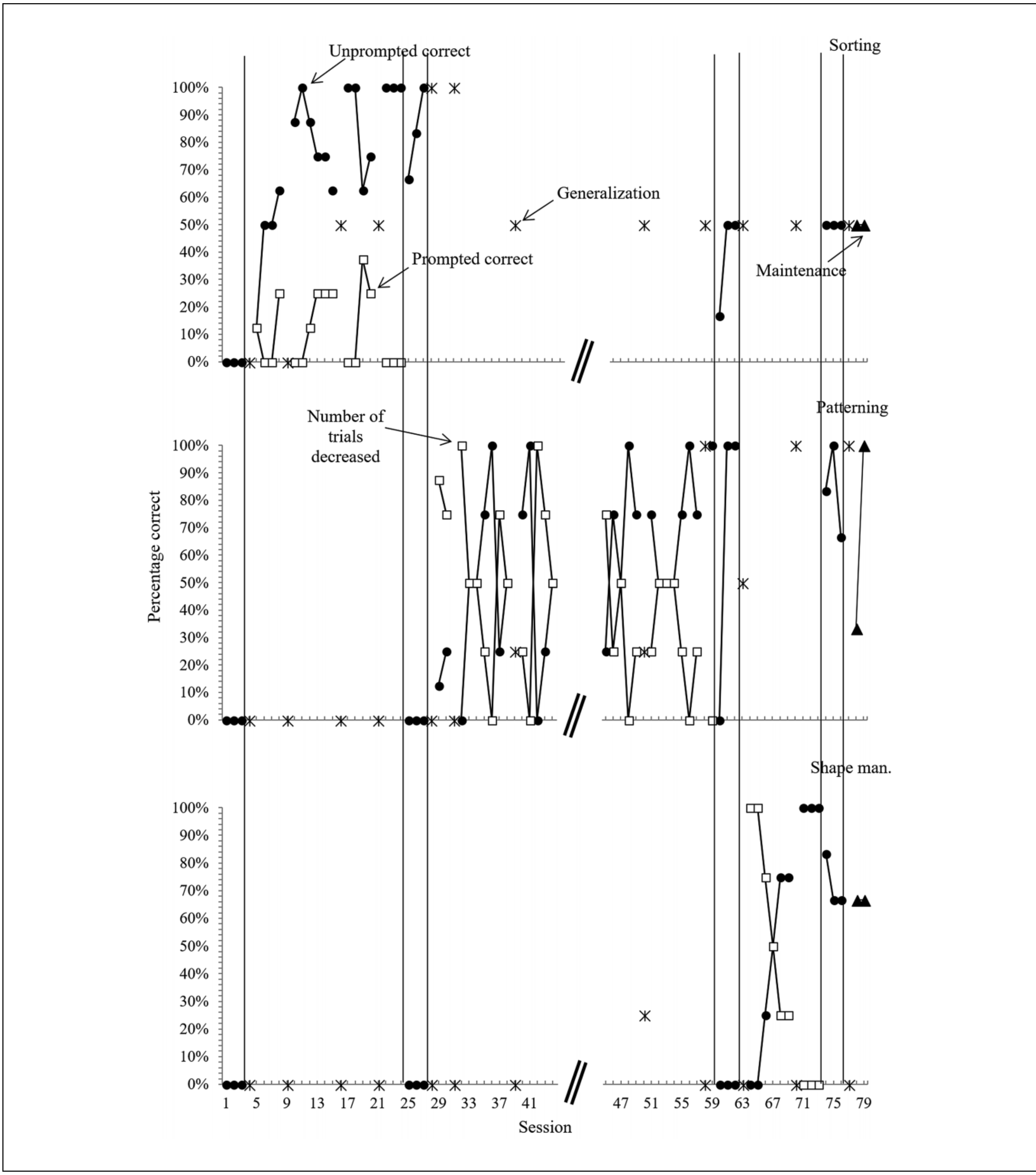

Figure 3. She'quan's acquisition, generalization, and maintenance of target behaviors.

generalization sessions in Tier 3 was again variable, but it reached 100% at the end of Tier 3.

In the first probe condition, She'quan demonstrated no shape manipulation behaviors in generalization sessions. In Tier 2, he demonstrated an average of 6.25% of shape manipulation behaviors, despite not having yet received instruction on them. In Tier 3, he demonstrated no shape manipulation behaviors in generalization sessions. She'quan never demonstrated shape manipulation behaviors in generalization sessions after intervention in Tier 3 was complete.

Maintenance of target behaviors. Data on She'quan's maintenance of target behaviors are presented in Figure 3. Maintenance probes are represented with a closed triangle. Across two maintenance sessions 1 month after intervention ended, She'quan demonstrated 50% of sorting behaviors, an average of 66.67% of patterning behaviors, and 66.67% of shape manipulation behaviors.

Dosage

Dosage was measured through the number of trials in each intervention condition. There was significant variability in the number of trials in each intervention condition both within and across participants. Jason required 48 trials to criterion in Tier 1 (sorting), 164 in Tier 2 (patterning), and 24 in Tier 3 (shape manipulation). She'quan required 128 trials to criterion in Tier 1 (sorting), 116 trials to near-criterion in Tier 2 (patterning), and 36 trials to criterion in Tier 3 (shape manipulation). Overall, patterning appeared to require a larger number of trials than other skills, and shape manipulation was acquired more quickly. There was significant variability across participants regarding the number of trials to mastery criterion for sorting.

Social Validity

Social validity data were collected both prior to and after the intervention. On both the pre- and postquestionnaires, the teacher strongly agreed that math skills are important for children to know and that children need instruction to learn math skills. This did not change from preintervention to postintervention. The teacher also responded to questions about how often she saw each target child engage in their target math skills, both prior to and after data collection. There was some change in the teacher's report of the frequency with which she observed each child demonstrating his target skills in the classroom, with a slightly higher frequency after intervention compared with prior to intervention for two of Jason's skills and one of She'quan's skills. The teacher also viewed a video of the instructional procedure and answered questions related to the procedure. She strongly agreed that the instructional procedure seemed appropriate to use with preschool children and that she could use the procedure in her classroom with both individual and small groups of children.

Discussion

The purpose of this study was to investigate whether a systematic instructional procedure was effective in helping children acquire, generalize, and maintain early math skills. A functional relation was demonstrated for both participants. Jason acquired all three of his target skills at criterion levels. She'quan acquired two of his target skills at criterion levels, and, although he did not reach mastery criterion for patterning, there was an effect. Criterion was three consecutive sessions at 100%; he had three sessions at 100%, but they were not consecutive. The generalization results were mixed. Jason demonstrated generalization of two skills and She'quan demonstrated generalization of one skill. The maintenance results also were mixed. One month after the end of the last probe condition, Jason maintained all three skills at or near 100%, and She'quan maintained all three skills at 50% to 60.66%.

The results of this study require further examination, particularly around the variability in the length of time it took children to acquire skills and the complexity of skills taught. Efficiency of learning was measured by examining the number of trials to criterion. When the number of trials in each tier decreases with each subsequent tier, this often suggests that children might be "learning to learn." That pattern was present in the current study only for She'quan. However, there was a great deal of variability within and among participants in the number of trials to criterion. Two factors might have contributed to this. First, over the course of the study, the number of trials per session was decreased due to session fatigue. This reduction in the number of trials appeared to prevent fatigue and was sufficient for children to acquire the behaviors taught. However, it is possible that presenting a greater number of trials per session would have resulted in more efficient acquisition. From the data, it is unclear how the change in number of trials per session affected the number of trials to criterion.

Second, the variation in the number of trials in intervention sessions, both within and across children, was also likely related to the complexity of the skill and the number of behaviors taught for each skill. Jason and She'quan had 164 and 116 trials to criterion (or near criterion), respectively, for the skill of patterning, in marked contrast to the number of trials to criterion Jason had for both of his other skills (48 for sorting, 24 for shape manipulation) and She'quan had for one of his other skills (36 for shape manipulation). These data suggest that patterning may have been a more difficult skill to learn or that duplicating patterns, rather than extending them, should have been taught. Duplicating patterns is considered a precursor to being able to extend patterns (Sarama & Clements, 2009), and it is possible that Jason and She'quan did not have this prerequisite skill. This highlights the need for sensitive measures of children's early math learning. Learning to pattern also may have been more difficult because multiple patterns were taught at the same time (AB and ABB). Thus, when selecting the number and type of behaviors to target for instruction for a given skill, attention must be paid to the complexity of the skill and the task demands associated with learning multiple examples of the skill at the same time. It is possible that children are able to generalize a skill more readily if they learn multiple behaviors (e.g., learning to complete AB and ABB patterns) and that this ameliorates concerns with how long it takes to acquire the skill. This is of a similar logic as training for generalization by using multiple exemplars (Stokes & Baer, 1977). It might also be that for more complex skills, it is simply necessary to teach one behavior at a time.

Contributions to the Literature

This study contributes to the literature on early math instruction for preschoolers in several ways. Evidence from this study indicates that children who are at risk, have not been exposed to high-quality math instruction, and have limited early math skills (based on assessment data) can acquire, and, to some extent, generalize and maintain, early math skills. The target skills and how they were selected in the current study is different than other studies on early math instruction. Much of the previous research focused on only those math skills related to number sense, and the identification of target skills was typically not individualized to the participant. Previous research rarely included measures of maintenance and generalization of math skills.

An additional contribution of this study is that the skills taught to children were based on an assessment of children's identified needs. The EMSI, which was developed as part of this study, was used to provide a measure of children's skills across all domains of early math. This tool potentially provided a more sensitive approach for selecting skills to target for instruction.

This study also provides evidence that teaching complex math skills to struggling learners may require multicomponent, individualized, and long-term interventions. All of these were necessary in the current study. This potentially limits the feasibility of such interventions.

Limitations of the Current Study

There were four main limitations to this study. The first is that the type of activity may have influenced the rate of acquisition of the target behaviors. The participants had more variable responding when art activities were chosen and they tended to make more initial errors in art activities than in manipulatives activities. It is possible that the art activities led to more mistakes because the children were focused on the creative aspect of the activity rather than attending to the task direction and math instruction. It is also possible that it was harder for children to correct mistakes made during the art activity (e.g., using the incorrect color marker when extending a pattern) or that the art component of the activity added additional demands to the activity, making it more difficult to complete (e.g., having to glue the shapes down after sorting them).

Jason chose the art activity for six of his Tier 2 sessions, and his Tier 2 data were variable, with 23 sessions needed to reach criterion. He chose the art activity only once in both his Tiers 1 and 3 intervention sessions, and he acquired those behaviors more efficiently (six and five sessions to criterion, respectively). Jason also made more initial errors in sessions in which art activities were used, compared with sessions in which manipulatives were used. He had 40% prompted corrects and prompted errors in art activities, compared with 12.79% when manipulatives were used.

She'quan chose art activities in two of his Tier 1 intervention sessions, and there was some variability in these data. He chose art activities in five of his Tier 2 intervention sessions, and these data also were variable, with 27 sessions to near-criterion. She'quan did not choose art activities in any of his Tier 3 intervention sessions, and there was little variability in these data. She'quan had 58.33% prompted corrects and prompted errors in art activities, compared with 31.97% when manipulatives were used.

The second limitation relates to the error correction procedure used. The initial error correction procedure allowed children to repeat incorrect responses. For example, for the behavior of sorting by size, if the child performed it incorrectly, the adult modeled the behavior again and had the child repeat the behavior independently. Thus, the child had a second opportunity to perform the behavior incorrectly. As a result, the error correction procedure was revised. If the child sorted by size incorrectly, the adult physically and/or verbally prompted the child to perform the behavior correctly, rather than expecting the child to perform the behavior correctly independently. Essentially, only unprompted correct and prompted correct responses were possible (unless the child refused to perform a behavior). It is possible that children's acquisition of target behaviors would have been more efficient had the revised error correction procedure been used from the beginning. This is because the revised procedure eliminated the possibility of the child making errors (without immediate help to correct the errors).

The third limitation relates to the generalization sessions. The goal was to use materials that were typically in the classroom to measure whether children could generalize the skills taught using researcher-provided materials to the existing classroom materials. However, that was not possible due to the lack of materials available in the classroom. In addition, the generalization materials for the Tier 3 skill (shape manipulation) for Jason and She'quan were problematic. Although it is possible that the children simply did not know the skill sufficiently well to generalize, another possible explanation relates to the stimuli used in the generalization setting. One stimulus appeared to be too easy (i.e., the outline of the picture suggested very clearly which shapes should be used to fill the picture), leading to both participants demonstrating the behavior before intervention in that tier. The other stimulus appeared to be too difficult (i.e., the picture was very complex, and it was difficult to determine how to put the shapes together to form the picture), leading to both participants being unable to demonstrate the behavior in the generalization setting even after reaching criterion in that tier.

The fourth limitation is that the social validity data were limited in scope and likely to be biased. Given that only one teacher completed the social validity questionnaire, it was not possible for it to be anonymous. In addition, the teacher knew the purpose of the study, which potentially impacted her answers to the questions.

Implications for Research

Additional research on this instructional procedure is needed. Further evidence is needed to demonstrate the effectiveness of the procedure in helping children acquire, generalize, and maintain math behaviors. It also is necessary to investigate the optimal number of instructional trials to provide to support efficient learning. In future research, additional exploration of embedding math instruction in art activities and play activities is warranted. Based on the results of the current study, embedding math instruction in art cannot be recommended.

Additional avenues of future research include investigating the effectiveness of the intervention: (a) when used with indigenous implementers (i.e., classroom staff); (b) with different populations of children, including children with disabilities; and (c) with a wider variety of math skills. Additional social validity evidence also is needed, both with a larger number of teachers and using other, more objective, measures, such as normative comparisons.

Implications for Practice

The primary implication for practice is that the systematic instructional procedure, when implemented with fidelity, can be effective in helping children acquire and maintain early math skills. However, further evidence is needed to establish that the instructional procedure is efficient. Long periods of instruction were required for at least some skills for both children in this study. This could be because the skills require this amount of instruction, but it also could be because the instructional procedure was not the most efficient approach for teaching some math skills or for some children.

Conclusion

This study provides initial evidence for the effectiveness of a systematic instructional procedure that can be used to increase young children's early math skills. However, additional research is needed to investigate this instructional procedure and determine its effectiveness for teaching early math skills to children with and without disabilities. Given the evidence regarding the power of early math skills to predict later academic achievement and the disparities among young children in early math skills, this research is of paramount importance.

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 was supported by the Early Childhood Special Education and Emotional Disturbance Leadership Training Grants (H325D070075 and H325D070083) awarded to Vanderbilt University by the Office of Special Education Programs, in the Office of Special Education and Rehabilitative Services, in the U.S. Department of Education.

References

Barton, E. E., Meadan-Kaplansky, H., & Ledford, J. R. (2018). Independent variables, fidelity, and social validity. In J. R. Ledford & D. L. Gast (Eds.), *Single-case research methodology: Applications and special education and behavioral sciences* (3rd ed., pp. 133–156). New York, NY: Routledge.

Charlesworth, R., & Lind, K. K. (2010). *Math and science for young children* (6th ed.). Belmont, CA: Wadsworth Cengage Learning.

Ciancio, D., Rojas, A. C., McMahon, K., & Pasnak, R. (2001). Teaching oddity and insertion to head start children: An economical cognitive intervention. *Journal of Applied Developmental Psychology*, *22*, 603–621.

Claessens, A., & Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers College Record*, *115*(6), 1–29.

Clements, D. H., & Sarama, J. (2007). *Building blocks—SRA real math, Grade PreK*. Columbus, OH: SRA/McGraw-Hill.

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

Clements, D. H., & Sarama, J. (2009). *Learning and teaching early math: The learning trajectories approach*. New York, NY: Routledge.

Clements, D. H., Sarama, J., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Maths Assessment. *Educational Psychology: An International Journal of Experimental Educational Psychology*, *28*, 457–482.

Clements, D. H., Sarama, J., & Wolfe, C. B. (2011). *TEAM—Tools for early assessment in mathematics*. Columbus, OH: McGraw-Hill Education.

Cross, C. T., Woods, T. A., & Schweingruber, H. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. Washington, DC: National Academies Press.

Daugherty, S., Grisham-Brown, J., & Hemmeter, M. L. (2001). The effects of embedded skill instruction on the acquisition of target and nontarget skills in preschoolers with developmental delays. *Topics in Early Childhood Special Education*, *21*, 213–221.

Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Division for Early Childhood. (2014). *DEC recommended practices in early intervention/early childhood special education 2014*. Retrieved from http://www.dec-sped.org/recommendedpractices

Duncan, G. J., Claessens, A., Huston, A. C., Pagani, L. S., Engel, M., Sexton, H., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446.

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D. C., . . . Pianta, R. C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly*, *25*, 177–193.

Gast, D. L., Lloyd, B. P., & Ledford, J. R. (2018). Multiple baseline and multiple probe designs. In J. R. Ledford & D. L. Gast (Eds.), *Single-case research methodology: Applications and special education and behavioral sciences* (3rd ed., pp. 239–282). New York, NY: Routledge.

Hardy, J. K. (2014). *Early Math Screening Instrument*. Unpublished instrument.

Holcombe, A., Wolery, M., & Werts, M. G. (1993). Effects of instructive feedback on future learning. *Journal of Behavioral Education*, *3*, 259–285.

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense mathematics achievement in first and third grades. *Learning and Individual Differences*, *20*, 82–88.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice*, *22*, 36–46.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*, 850–867.

Kidd, J. K., Curby, T. W., Boyer, C. E., Gadzichowski, K. M., Gallington, D. A., Machado, J. A., & Pasnak, R. (2012). Benefits of an intervention focused on oddity and seriation. *Early Education and Development*, *23*, 900–918.

Klein, A., & Starkey, P. (2002). *Pre-K mathematics curriculum*. Glenview, IL: Scott Foresman.

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness*, *1*, 155–178.

Ledford, J. R., Lane, J. D., & Gast, D. L. (2018). Dependent variables, measurement, and reliability. In J. R. Ledford & D. L. Gast (Eds.), *Single-case research methodology: Applications and special education and behavioral sciences* (3rd ed., pp. 97–132). New York, NY: Routledge.

Markowitz, J., Carlson, E., Frey, W., Riley, J., Shimshak, A., Heinzen, H., . . . Lee, H. (2006). *Preschoolers' characteristics, services, and results: Wave 1 overview report from the Pre-Elementary Education Longitudinal Study (PEELS)*. Rockville, MD: Westat.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Who is at risk for persistent mathematics difficulties in the United States? *Journal of Learning Disabilities*, *49*, 305–319.

Mullen, E. M. (1995). *Mullen Scales of Early Learning*. Bloomington, MN: Pearson.

Ochsendorf, R. (2016). Advancing understanding of mathematics development and intervention: Findings from NCSER-funded efficacy studies. *Journal of Research on Educational Effectiveness*, *9*, 570–576.

Office of Head Start. (2015). *The head start early learning outcomes: Ages birth to five*. Arlington, VA: Office of Head Start, Administration for children and Families, U.S. Department of Health and Human Services.

Pasnak, R., Greene, M. S., Ferguson, E. O., & Levit, K. (2006). Applying principles of development to help at-risk preschoolers develop numeracy. *Journal of Psychology: Interdisciplinary and Applied*, *140*, 155–173.

Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*, 375–394.

Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology*, *32*, 146–159.

Ramani, G. B., Siegler, R. S., & Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of Educational Psychology*, *104*, 661–672.

Repp, A. C., Deitz, D. E. D., Boles, S. B., Deitz, S. M., & Repp, C. F. (1976). Differences among common methods for calculating interobserver agreement. *Journal of Applied Behavior Analysis*, *9*, 109–113.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. New York, NY: Routledge.

Schiller, P., Flor Ada, A., Campoy, I., & Mowry, B. (2010). *Frog Street Pre-K*. Crandall, TX: Frog Street Press.

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, *11*, 655–661.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—Improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*, 545–560.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, *10*, 349–367.

Wolf, M. M. (1978). Social validity: The case for subjective measurement of how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, *11*, 203–214.