Development of Curricula, Teacher Supports, and Assessments
for Pre-Kindergarten Mathematics and Science

Mable B. Kinzie, Ph.D., Robert C. Pianta, Ph.D., Carolyn R. Kilday, B. A.,
Patrick R. McGuire, M.Ed., & Ashley M. Pinkham, M.A.

Center for Advanced Study of Teaching & Learning (CASTL)
University of Virginia

Paper presented on March 2, 2009 at the annual meeting of the
Society for Research on Educational Effectiveness (SREE), Washington, DC.

Correspondence may be addressed to the primary author, at:

MyTeachingPartner-Mathematics/Science, CASTL, University of Virginia,
P.O. Box 400878, Charlottesville, VA 22904-4878. E-mail: kinzie@virginia.edu

The research reported here was supported by the Institute of Education Sciences, U.S.
Department of Education, through Grant R305B040049 to the University of Virginia.
The opinions expressed are those of the authors and do not represent views of the U.S.
Department of Education.
Development of Curricula, Teacher Supports, and Assessments for Pre-Kindergarten Mathematics and Science

Background/context:

Need for Early Childhood Curricula in Mathematics and Science. Informal mathematical knowledge undergoes considerable development during the preschool years (Baroody, 1992; Beilin & Klein, 1982; Cooper, 1984; Mix, 2002; Newcombe & Huttenlocher, 2000; Starkey & Cooper, 1995; Wynn, 1990; Zur & Gelman, 2004), and lays the foundation for the development of formal mathematical knowledge and skill in elementary school (Geary, 1994; Ginsburg, 1998). The development of this knowledge can be fairly easily nurtured through both play and instruction. This is particularly important for children who live in poverty, as they have been shown to experience considerable difficulty in mathematics and early intervention programs can address equity issues and narrow the performance gap (Clements, 2001).

Young children also develop substantial informal science knowledge, by actively engaging with their environments to understand observed phenomena and develop essential process skills (Eshach & Fried, 2005; Gallenstein, 2003; Lind, 1999; Platz, 2004). These skills, along with conceptual understandings and inquiry strategies, begin to develop as early as infancy, with the sophistication of children’s competency developing with age (Klein, 1998; Lind, 1999; Meyer, Wardrop, & Hastings, 1992; Piaget & Inhelder, 2000). Environmental effects are important—the lack of needed stimuli may result in a child’s development not reaching its full potential (Hadzigeorgiou, 2002).

Basis for Curricular Design. The MTP-Math/Science curricula specifically target the teaching and learning of children at risk of early school failure, a population for whom achievement gaps in mathematics and science are visible even in Pre-K years. MTP-Math is based on Focal Areas defined by the NCTM (2006) for Pre-K through the 8th grade and developmental trajectories for Mathematics from Pre-K to grade two advanced by Clements (2004), and further focused through a review of state Pre-Kindergarten standards. The domains of MTP-Math include: Number, Operations, Geometry, and Measurement. Within the Science domain, the AAAS K-12 Science Benchmarks (1993) identified conceptual and skill domains targeted for Kindergarten and beyond; state standards helped refine our curricular focus for Pre-K within these domains. The MTP-Science domains include the Life, Earth, and Physical sciences.

To provide authentic points of inquiry, our year-long curricular trajectories reflect seasonal changes; we used these trajectories to help extend children’s thinking across the year. MTP-Math/Science offers a variety of inquiry-based activities, with ample opportunities for children to observe, predict, collect, analyze and communicate both their processes and results. We emphasize a balanced integration of student-centered, highly-contextualized and meaningful interactions with teacher-directed, scaffolded target exposures to key concepts. The work of Ginsburg (Ginsburg & Golbeck, 2004) has been influential, as we emphasize opportunities to encourage children’s thinking and model/elicit mathematics and science language to express that thinking. Children are encouraged to Engage, Investigate, Discuss, and Extend, a modification of the 5E Model (Bybee, et al., 2006). We use children’s literature to anchor investigations, providing background context and comfortable entry points, further supporting the development of literacy and language. Design of MTP-Math/Science is illustrated in Fig. 1.
Development of Year-long Curricular Trajectories. Throughout the course of the academic year (September – May) we deliver a total of 66 activities in Math, and another 66 activities in Science. In addition, across the 33 weeks teachers and students engage in center-based activities that help to deepen children’s understandings in both Math and Science. Table 1 demonstrates how the MTP-Math curricula address a specific skill (comparing and describing objects with respect to length) within the Math domain of measurement. Collectively, the activities offer the opportunity to revisit and deepen related measurement skills and to build upon students’ previous conceptual understanding.

Iterative Curricular Development. An iterative rapid-prototyping approach helps designers ensure that products are effective and engaging. Our activity design began with creative brainstorming and continued to development of early prototypes that were repeatedly revised, evaluated by teachers and tried out in classrooms before they were “finished.” Prior opportunities to learn help determine what is developmentally appropriate (National Research Council, 2007), and our year-one classroom observations (in three pilot classrooms) suggested the importance of multiple opportunities for students to engage with targeted concepts. These opportunities are now achieved using weekly Math and Science Center experiences, strengthening the connections between activities throughout the year. We also observed teachers’ tendency to focus on the doing of activities rather than on encouraging students’ thinking about them. In response, we have placed additional curricular emphasis on opportunities for modeling language and eliciting thought.
<table>
<thead>
<tr>
<th>Activity</th>
<th>Objective</th>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept Wk 3 Math Counts – Shapes</td>
<td>Compare 2 objects in length</td>
<td>Students explore properties of circles and squares and compare the lengths of the sides of various shapes. Provides a basic introduction to non-standardized measurement.</td>
</tr>
<tr>
<td>Oct Wk 1 Building Triangles and Rectangles</td>
<td>Compare and describe the relative length (longer, shorter, same) of 2 objects</td>
<td>Students advance to shape building, rather than simply describing and identifying, using manipulatives to build and compare rectangles and triangles. Language used to describe relative length.</td>
</tr>
<tr>
<td>Nov Wk 1 Ordering by Length</td>
<td>Compare and describe the relative length (longest, shortest, middle, same) of 3 objects</td>
<td>Students use objects throughout the classroom to compare lengths and place them in the correct order. This activity introduces a third object to be compared, and challenges students to recall their prior measurement skills.</td>
</tr>
<tr>
<td>Jan Wk 2 Measuring With Linking Cubes</td>
<td>Measure the length of 1 object using other standardized objects</td>
<td>Following the reading of <em>Inch by Inch</em>, by Leo Lionni, students practice measuring objects using standardized objects (linking cubes). This activity serves as the first introduction to standardized measurement, and emphasizes the importance of using the same starting point when making measurement comparisons.</td>
</tr>
<tr>
<td>Feb Wk 2 Monkey Measurement</td>
<td>Measure the length of 2 objects using other standardized objects</td>
<td>Students construct their own monkey rulers (10 units long) and begin to conduct standardized measurements of multiple objects. Students are now incorporating all of the previous non-standardized and standardized measurement skills, as well as using all of the language acquired throughout the year.</td>
</tr>
<tr>
<td>Apr Wk 1 Length Comparison</td>
<td>Compare and order 5 objects in length</td>
<td>The activity revisits and builds upon the concepts presented in Nov Wk 1. Students are responsible for measuring a variety of animals (on picture cards) and placing them in order by length. Emphasis is placed on different ways to order, as well as ordinal language such as first, second, third, etc.</td>
</tr>
<tr>
<td>May Wk 4 Monkey Measurement Part II</td>
<td>Measure the length of 2 objects using other standardized objects</td>
<td>The culminating measurement activity for the year. Students use the same monkey rulers (from Feb Wk 2) but extend the length of the rulers to 15 units. Throughout the activity, students use comparative language, ordinal vocabulary, and make predictions about the lengths of objects before conducting actual measurements.</td>
</tr>
</tbody>
</table>
Teacher feedback helped us to pare down our activity instructions to the most important elements and we revised our graphical layout. The resulting layout allows the entire activity to be easily reviewed on facing pages (Figure 2 shows the evolution of the same activity from years one to two). Every activity follows the same uniform framework:

- **Engage**
  - Math or Science Chant
  - Inquiry-based question or prompt
- **Investigate**
  - Explores the primary learning objectives
- **Discuss**
  - Presents students with a series of questions intended to reflect on the activity
  - Emphasis is placed on language modeling and explanation of activity procedures (how kids think)
- **Extend**
  - Activity concludes with an *optional* set of extension ideas

![Figure 2. Evolution in Activity Layout](image)

The Need for Pre-Kindergarten Teacher Development in Mathematics and Science. Even when offered validated curricula, teachers in Pre-K programs are highly likely to *not* implement them with high quality or fidelity, largely as a function of their lack of content knowledge and lack of confidence, in addition to general struggles to conduct instruction effectively (Pianta et al, 2005). The preschool teaching force has been “generally characterized by inadequate preparation and offered inadequate ongoing professional development and these teachers generally have few years of experience because of high turnover rates” (National Research Council [NRC], 2005, p 16). Teachers in early childhood programs lack confidence in their knowledge of science and science education pedagogy (Fensham, 1991; Garbett, 2003).
same tends to be true in mathematics: most preschool teachers who do not have appropriate and effective training in mathematics do not value mathematics as a priority for young children, and they lack confidence in their own abilities to teach mathematics (Baroody & Coslick, 1998). Well-designed professional development experiences have enhanced teachers’ learning of concepts in mathematics and science and their teaching practices (Katz, 1999) and have lead to improved preschool program quality (National Research Council [NRC], 2001).

**Teacher Supports and the MTP Logic Model.** In their review of primary prevention programs for mental disorders, Greenberg and colleagues (Greenberg, Domitrovich & Bumbarger, 2001) found a significant relationship between quality of implementation and classroom outcomes. Large-scale observations conducted for the Multi-state Study of Pre-kindergarten and the Statewide Early Education Programs (SWEEP; LoCasale-Crouch et al., 2007; Mashburn et al., 2007; Pianta, Howes, Early, Clifford, Bryant, & Burchinal, 2003) indicate that variation in instructional and emotional support quality were directly related to growth in children’s achievement test scores and social behavior ratings. These findings underline the importance of professional development emphasizing and assessing both high fidelity and high quality.

In our earlier research with the MTP model to support language and literacy (Pianta, Mashburn, Luckner, Myers, & Kilday, 2008), the MTP teacher professional development program improved the quality of classroom interactions that Pre-K students experience, which in turn, promoted children’s development of language and literacy skills. Targeting teachers’ interactions with students as the focus of professional development and training may be particularly beneficial because these interactions are the proximal mechanism responsible for effects on children’s early experiences. This forms the basis of the logic model for MTP-Math/Science, depicted in Figure 3.

![Figure 3. MTP-Math/Science Logic Model](image-url)
**Design of Teacher Supports.** One of the primary ways that we are promoting high quality interactions between students and teachers is through the use of a range of supports, some embedded in the activities themselves and others accompanying the curricula on-line. Every activity offers *Make It Work*—recommendations for differentiating the instruction for children who are more or less developed in the knowledge and skills.

On our website, we provide teachers with access to additional supports, to review *before* implementing an activity in their own classroom. Brief Teaching tips in every activity emphasize *How Kids Think* (common ways of understanding; frequent misconceptions), *Be the Best* (suggested teaching strategy particularly important for the activity), or *The Big Idea* (underlying important concept, fact, application).

Also embedded in every activity are Demonstration Videos of two-three minutes in length. These videos show Pre-K teachers implementing the activity with high levels of quality and fidelity. They show some of the ways the activity can best work to encourage children’s knowledge and skill development, and help teachers plan how they are going to effectively implement the activity. The classrooms and students depicted in the videos reflect the community of practice for at-risk Pre-Kindergarten programs.

![Figure 4. Demonstration Video: Measuring Body Parts](image-url)

Every week, a different Quality Challenge emphasizes a single aspect of high quality teaching, drawn from the constructs of the Classroom Assessment Scoring System (CLASS; Pianta, LaParo, & Hamre, 2008). The CLASS is comprised of three domains: emotional support, classroom organization, and instructional support, each of which is split into three to four dimensions. The weekly Quality Challenges specifically relate to the curricular activities undertaken that week, and alternate between Math and Science activity settings. Each Challenge provides the teacher with a prompt for reflection, prior to watching a short (less than two minutes) video clip of teaching practice. After viewing the clip and formulating their response,
teachers can compare their reactions to that of a Math or Science Education expert. In the process, teachers reflect on the practice of others and think about their own instructional practice.

Our goal in the design of all of these teaching supports has been to produce teaching aids that would be quick to use and offer substantial value to teachers. For teachers desiring more, we offer access to a rich library of Quality Teaching videos, depicting and describing practices from a range of settings and curricular areas (Language, Literacy, and Social-Emotional Relationships, in addition to Math and Science).

Purpose/objective/research question/focus of study:

In year two of our project, we are applying the MTP model to the design and development of embedded on-line teacher supports that emphasize high quality classroom interactions and high fidelity implementations. We are also continuing to iteratively revise/evaluate the curricula. Our development focus is:

- How might we best support instructional practice with on-line, embedded teacher supports?

We are also considering the following research questions:

- Are classroom quality and curricular fidelity positively correlated with pre/post gains in children’s mathematics learning? …with end-of year scores in science?
- Are there correlations between quality and fidelity?
- As the amount of teacher exposure to the curricula and embedded math and science concepts increases, do teaching quality and fidelity also increase?

Setting & Population:

We are currently implementing the MTP-Math/Science curricula in eight Pre-K classrooms from a state-wide initiative providing publicly-funded Pre-Kindergarten to children who have one or more risk factors for later school failure (poverty, second language learners, or health or developmental problems). These classrooms represent a convenience sample: teachers from several school districts were approached and invited to help evaluate our curricula. Teachers are provided with a small honorarium for their participation, in addition to receiving the curricula and required teaching manipulatives. All classrooms are located in a public school buildings and not a part of Head Start. Two classrooms had 16 children, four had 17 children, and two had 18 children. Across all eight classrooms, there were 63 girls and 73 boys. In terms of age, students were between 42 and 60 months old; all will be eligible for kindergarten in the 2009-2010 school year. In order to make simple comparisons, we examined the performance of 16 children drawn from two non-participating classrooms. Both serve four-year old children and have no more than 16 students in the class; one of these classrooms was an additional state-funded classroom and the other was federally-funded to serve the same at-risk population.

Intervention/Program/Practice:

Teachers implement MTP-Math/Science curricular activities six times/week (two math and two science activities, with follow-up math and science centers). Activities range from 10-25 minutes in length; the majority are facilitated in small group settings (six to eight students) while
the remaining activities take place in whole group format. Center activities are intended to help students re-visit key concepts or ideas, and provide for additional exploration or practice. To help us evaluate the curricula and examine effects on teaching practice, teachers videotape their implementation of all MTP activities.

Teachers are also being exposed to a limited number of on-line teacher supports during this year of the project, as we develop and test prototypes for these features. These supports include video demonstrations of high quality, high fidelity teaching practice and interactive “Quality Challenges,” among other elements. Most of these are embedded within the curricula, to be encountered in the context of teaching practice.

Research Design:

Our current research is quasi-experimental, allowing implementation of our curricula and consideration of the relationships between variables.

Data Collection and Analysis:

Evaluation Analyses. We are conducting video observations of all activities across the year. To help us ensure that a range of teaching practice is being observed, but to control the observation time required, we have separated teachers into higher/lower quality groups, based on fall application of the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, and Hamre, 2008). To evaluate each activity, we randomly select a video from one teacher in the higher quality group and one from the lower quality group (selection without replacement until all teachers have been observed). We also obtain useful information from periodic focus groups, the end-of-year teacher survey, and end-of-year teacher interviews.

Student Characteristics. Parents or guardians of all children in the treatment classrooms were asked to fill out demographic questionnaires about their preschooler, and 98 agreed (response rate of 72%). Children in this group were evenly distributed in terms of gender (boys = 47, girls = 51). Only 13 parents reported that their child has a diagnosed disability, and English is the predominant language spoken at home in 93% of the sample. In reporting the ethnicity of their preschooler, 49% reported their child as White/Caucasian, 29% as Black/African American, and the remainder were reported as another race or a mix of racial backgrounds. Just over half of the respondents (52%) reported earning an annual income of $35,000 or less. When asked about the highest level of education attained by the mother/female guardian, only 12% of the sample reported having a bachelor’s degree or above; 84% reported having finished high school. From the 16-18 children in each classroom, we randomly selected 8 for participation in direct assessments (we eliminated from participation any student for whom the primary language was not English as well as any students with Individualized Educational Plans for anything but speech). The demographics of the children selected for assessment (n=47) did not differ significantly from the overall classroom. Eight children were also selected from each of our two control classrooms, with a total of 5 boys and 11 girls in the control sample.

Use teacher roster to fill out the numbers below (for example, number 3 on the teacher roster gets placed next to number 3 on the list of random numbers). Any students with an IEP (other than speech), who are ESL or who are not present on the day of assessment are excluded from the
pool (crossed off the list). The first eight students who are not crossed off the random number list below are selected for assessment.

**Teacher Characteristics.** The average age of the teachers in the study is 42 years. The majority of teachers are white \((n = 6)\), and the remaining two are African American. They have been working with preschool children for an average of 3.8 years each (range = 0 to 9 years). The teachers had an average of 5.9 years \((SD = 9.2)\) working with children in kindergarten and 2.3 years \((SD = 3.1)\) working with children older than kindergarten. One teacher also reported working with children younger than pre-k for 5 years.

In terms of education, all of the teachers hold a bachelor’s degree and are specifically certified to teach 4-year-old children. In addition, three teachers have one year of graduate coursework, and three others have a master’s degree. Teacher self-efficacy was assessed via the *Beliefs About Teaching* scale, which indicates the degree to which a teacher feels she can effect change in her students. The scale ranges from a low of 12 to a high of 108; teachers in our sample evidenced a moderately high sense of self-efficacy (range = 70-93). To determine the extent to which the teacher adhered to more authoritarian views of child development, the *Ideas about Children* scale was given. Scale scores range from 16 (believe that learning should be child-driven) to 80 (teacher’s role is authoritarian). Participating teachers returned scale scores ranging from 28 to 55 with a mean of 40.4 \((SD=10.4)\), suggesting teachers believed in a balance between teacher authority and child-driven experiences.

Six of the eight reported taking at least one course in the teaching of math and an additional course in the teaching of science. However, only half reported participating in supervised student teaching of any math and science lessons. Five of the teachers reported taking a college algebra, trigonometry, or elementary functions class and five reported taking a statistics and probability course. One teacher reported having taken a college geometry course. In terms of science one teacher reported taking a chemistry course, two reported taking earth science, one reported taking astronomy, three reported taking geology, one reported taking physical science, four reported taking biology or life sciences, and one teacher reported taking physics.

Within the past three years, only limited amounts of professional development in math and science were reported: Three teachers did not report any math- or science-specific professional development while among the other five, the following activities were reported: a post-graduate university course in the teaching of math, and attendance at a conference on mathematics teaching (one teacher), peer observation and mentoring for math and science teaching (one teacher), regular meetings to discuss issues related to the teaching of math and science (three teachers), and workshops on the teaching of math (four teachers) and science (two teachers)

In terms of instructional practices, five of the eight teachers reported engaging in daily mathematics activity, with one indicating math activity 3-4 times a week, and two engaging in math once or twice each week. Seven of the eight reported spending 30 minutes or less on math each time, with the eighth spending between 31 and 60 minutes each time. Where science activities were concerned, five of the eight teachers reported engaging in science one to two times a week, two reported three to four times per week, and one teacher reported daily science activities. All eight teachers reported that the science activities took less than thirty minutes to complete. Seven of the eight classrooms had math centers and five had science or nature centers.
For a point of comparison, all eight teachers reported daily reading and language arts instruction, spending between 1 and 30 minutes (four teachers), 31 and 60 minutes (three teachers), and 61-90 minutes (one teacher), and all eight classrooms had reading centers.

Technology usage and comfort may affect the degree to which our online teacher supports will be utilized. While technology usage may vary from district to district, teachers in this sample reported frequent use of the Internet ($M = 4.5$, $SD = 8.0$, range = 1 [never] to 5 [every day]) and e-mail ($M = 5$, $SD = 0.0$; same range). Five of the eight reported feeling very comfortable with using the web and using e-mail, while the other three reported being somewhat comfortable with both.

Child Outcomes: Mathematics. Child outcomes in mathematics are being determined through scores on two measures. The Test of Early Mathematic Ability – 3rd Edition (TEMA-3; Ginsburg & Baroody, 2003) reflects a child's knowledge of both formal and informal mathematic abilities, focusing on the domains of counting, one-to-one correspondence, numeral recognition, number facts, calculation, and understanding of concepts. The Early Mathematics Assessment – Geometry (EMA-G; Clements & Sarama, 2008) covers geometry and measurement, including shape identification, matching, and production, pattern identification, reproduction and extension, length and weight. Because we must limit the time spent in assessments to 20-30 minutes for each measure, we are operating with a subset of items taken from EMA-G, together with items we have adapted or developed to ensure we assess all aspects of geometry and measurement indicated in our curricular goals.

Child Outcomes: Science. There are no currently available, valid, reliable assessments focused on pre-K students’ understanding of specific science content. Some assessments of general cognitive abilities (e.g., Woodcock-Johnson) or vocabulary (e.g., Peabody Picture Vocabulary Test) include a limited set of scientific concepts. However, these assessments fail to include important elements such as making observations and conducting investigations. We are currently developing a Pre-K science assessment to indicate children’s understanding as a function of the MTP-Science curriculum and to measure children’s ability to integrate and generalize those concepts. The science assessment will be comprised of two components: life sciences and physical/earth sciences.

Assessment items have been based on curricular objectives and tasks adapted from a variety of published studies. In accordance with the National Science Education Standards, each concept is assessed in two or more ways. Multiple methods of assessment provide converging evidence regarding children’s understanding of specific concepts, increasing reliability of the measure and indicating the flexibility of children’s knowledge. Multiple assessment methods also help identify possible variability in children’s engagement and understanding of the particular items, and reduce the possibility of drawing erroneous conclusions on the basis of individual characteristics. In our setting, children’s ability to respond to open-ended verbal narrative questions may be a function of verbal ability or shyness in addition to their concept knowledge or inquiry skills. We are therefore supplementing open-ended questions with forced choice identification tasks, to which children can respond via pointing, sorting photographs, etc., depending on the task. Including items of both of these types will also indicate the depth of children’s knowledge, as free recall tasks are invariably more difficult than recognition memory tasks (e.g., Anderson & Bower, 1972, 1974). Table 2 provides sample assessment items of both types for life, earth, and physical sciences.
## Table 2
Sample of Multiple Methods Used for Science Direct Assessment

### Domain: Life Sciences
**Concept:** Living and Non-Living Things

<table>
<thead>
<tr>
<th>Identification</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Card sort: Ask child to place living things in one box, non-living things in a separate box</td>
<td>• Describing features of living and non-living things</td>
</tr>
<tr>
<td>– 10 living entities: 5 animates (i.e., animals), 5 inanimates (i.e., plants)</td>
<td>– 4 pairs of photos: teddy bear vs. grizzly bear; bird vs. airplane; snake vs. train; dog vs. plant</td>
</tr>
<tr>
<td>– 10 non-living entities: 5 artifacts (e.g., car), 5 natural (e.g., mountain)</td>
<td>– “How is this one is different from this one?”</td>
</tr>
<tr>
<td>– Coding: Accuracy</td>
<td>– “How is this one is the same as this one?”</td>
</tr>
<tr>
<td></td>
<td>– Coding: Content of response(s)</td>
</tr>
</tbody>
</table>

### Domain: Earth Sciences
**Concept:** Day and Night (Celestial Objects)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Identifying celestial objects</td>
<td>• Describing features of the sky</td>
</tr>
<tr>
<td>– Show picture of daytime sky and nighttime sky</td>
<td>– Sets of pictures show changes in the sky: sunrise to sunset; new moon to full moon; clouds moving across sky</td>
</tr>
<tr>
<td>– “Point to the picture that shows daytime [nighttime].”</td>
<td>– “Look at these pictures. What do you see in the pictures?”</td>
</tr>
<tr>
<td>– “Point to the sun [clouds, moon].”</td>
<td>– “What’s happening in the pictures?”</td>
</tr>
<tr>
<td>– Coding: Accuracy</td>
<td>– Coding: Content of response(s)</td>
</tr>
</tbody>
</table>

### Domain: Physical Sciences
**Concept:** Liquids and Solids

<table>
<thead>
<tr>
<th>Identification</th>
<th>Narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Identify liquids and solids</td>
<td>• Predicting changes in liquids and solids</td>
</tr>
<tr>
<td>– Show child a container of water and a block</td>
<td>– Show child a distinctively-shaped container</td>
</tr>
<tr>
<td>– “Point to one that is a liquid [solid].”</td>
<td>– “Let’s say we put the water [plastic block] in this container. What would the water [plastic block] look like?”</td>
</tr>
<tr>
<td>– Coding: Accuracy</td>
<td>– “How do you know?”</td>
</tr>
<tr>
<td></td>
<td>– Coding: Content of response(s)</td>
</tr>
</tbody>
</table>
To ensure the clarity of the questions and to determine the appropriate length of the assessments overall (young children have limited attention spans and cognitive resources), we are engaging in two cycles of pilot testing and revision, with the goal of clearly understood questions and an assessment length of 20-30 minutes for each of the science measures. The pilot testing is being conducted with a convenience sample of 12 consented children from a private preschool with a teacher:child ratio that is comparable to the Pre-Kindergarten classrooms serving our target population. Assuming that these children were from a more privileged background than our targeted pre-K children and assuming all of the advantages associated with higher SES, such as larger vocabularies, we felt confident in needing to revise or eliminate any items on the assessments that these children could not interpret. Likewise, with higher SES related to school readiness, including the ability to stay focused on a particular task, if our assessments ran too long for these children, or a particular task was not engaging to them, we felt confident in needing to revise the assessments before giving them to the pre-K population.

During pilot testing, children provided responses for all forced-choice items. They did so without significant hesitancy or needing to ask for clarification. Regardless of the accuracy of their responses, their responses indicate that they understood the questions being asked.

By contrast, there were a number of open-ended questions that were not answered. It was not uncommon for children to respond with “I don’t know.” Young children are sensitive to feelings of uncertainty (Fay & Klahr, 1996; Robinson & Whittaker, 1985); as a result, they may be hesitant to provide an answer when they are not confident in their responses (Pillow, 2002). This may particularly be the case for open-ended questions because the questions’ non-dichotomous nature may highlight children’s feelings of uncertainty, particularly as the children monitor the relative frequency of perceived errors over time (Pasquini, Corriveau, Koenig & Harris, 2007). It is possible that children understood the questions and may have even had some conceptual understanding, but were hesitant to venture guesses. “I don’t know” and non-responses did not appear to be more/less likely to occur based upon topic, etc. For instance, one child provided explanations for how birds are different than airplanes, but responded “I don’t know” to how snakes are different than trains. This suggests that the child has some understanding of living versus non-living entities, but that she was unable to generalize that understanding to all contexts. Open-ended questions were frequently “incomplete” (from a coding perspective). For example, consider the item “What do animals need to grow?” There are four important answers: food, water, light, and air. On average, children spontaneously offered one response. With prompting, some children provided additional responses. Without sufficient prompting from the test administrator, children’s knowledge on open-ended questions may appear more limited than it actually is. Our protocol therefore provides such prompting for items with multiple correct responses (e.g., “Can you think of anything else?”) or instances in which children do not spontaneously provide explanations or justifications for their responses (e.g., “How do you know?”).

Our observations of children’s performance led us to retain, revise, or discard assessment items. For example:

- Children readily sorted the photographs of living and non-living things. Although their performance reflected common biases, such as categorizing living things according to animacy (e.g., plants are not alive, whereas cars and bicycles are alive), we could determine that the task was at the appropriate level irrespective of children’s actual performance.
Children readily used modality-specific sensory information to identify object properties (e.g., color, scent, flavor). However, they appeared confused when asked to describe how they identified the property, often erroneously showing through actions how they found out. O’Neill and Chong (2001) reported that 4-year-olds may be experiencing a transition in their ability to recognize the origins of their modality-specific knowledge. Given this, we revised the assessment item to explicitly highlight the requested response type (i.e., emphasize verbal rather than action responses).

Children appeared particularly confused on some of the items assessing life cycles. For instance, none of the children tested were able to explain the egg-tadpole-toad transition. It was unclear whether children did not understand the transition, did not understand what was being asked of them, or could not tell what was being depicted in the photographs. Given this, we eliminated this particular item while retaining items on egg-chick-chicken, caterpillar-cocoon-butterfly, and seed-seedling-sunflower transitions.

We are still in the process of revising some items, developing an administration manual, and testing administrators for consistency in procedure. The resulting measures for life and physical/earth science will be employed in an end-of-year assessment to children in our participating Pre-K classrooms, in addition to those from two control classrooms. In addition to on-line (live and in person) coding, we plan to videotape these sessions to fully code children’s narrative responses off-line. With this administration we will be able to ascertain more precisely whether the measures are at the appropriate level of difficulty for the Pre-Kindergarten students.

Child Outcomes: Receptive Vocabulary. The Peabody Picture Vocabulary Test –III (PPVT-3; Dunn & Dunn, 1997) is a test of receptive vocabulary for ages two years, six months up to adulthood. For each word provided, respondents are asked to select a corresponding picture from a set of four. The receptive language data will be used as a covariate when considering other child outcomes.

Teacher Ratings: Mathematics and Science. Teachers will rate children’s mathematics skills, using the seven-item Academic Rating Scale (ARS) for Mathematics (developed for the ECLS-K, National Center for Education Statistics, 2002), including five additional items to address missing aspects of geometry and operations (for example, “Demonstrates an understanding of geometric shapes - for example, sorting, naming, and describing the characteristics of circles, squares, rectangles, and triangles.”) For teacher rating of science knowledge and skills, three items have been drawn from the ARS General Knowledge (also developed for the ECLS-K); additional items will address missing concepts and skills. Skills on these measures are evaluated against a five-point scale: 1 (Not Yet Proficient), 2 (Beginning), 3 (In Progress), 4 (Intermediate) and 5 (Proficient).

Implementation Outcomes: Dosage. Teachers will complete monthly questionnaires indicating which curricular activities they have completed; these will be cross-checked against the videotapes they submit (teachers videotape all curricular activities). Teachers will indicate via end-of-year survey the time they spent facilitating curricular activities each week.

Implementation Outcomes: Fidelity. Based on the design theories undergirding our curricular design, and drawing on previous measures of fidelity in early childhood curricula (Hamre & Pianta; Ertle & Ginsburg, 2006; Clements & Sarama, in press), we developed a rating
scale for implementation fidelity. Primary constructs measured include: Activity Completion, Materials & Environment, Teacher's Instruction, Ensuring Engagement, Content Coverage, and Supporting Cognition & Language. Scale items address classroom conditions as well as teacher and child behaviors. This assessment has been iteratively tested and revised, most recently with a team of five coders jointly reviewing and coding randomly selected videos of a teacher implementing a math or science lesson. At this point, only minor item revisions are being made, to better operationalize definitions and provide examples for each item to be included in a coding manual.

Implementation Outcomes: Quality. The Classroom Assessment Scoring System (CLASS) for Pre-Kindergarten was developed by Pianta, La Paro, and Hamre (2008) to examine the interactions between children and adults in classrooms that lead to better child outcomes. It was based on large-scale classroom observation studies in the NICHD Study of Early Child Care (NICHD ECCRN, 2002; Pianta et al., 2002) and the NCEDL Multi-State Pre-K Study (Bryant et al., 2002). The CLASS is comprised of three domains: emotional support, classroom organization, and instructional support, each of which is split into three to four dimensions. Each dimension is scored based on a range of behavioral indicators.

For both Fidelity and Quality, we will randomly select two videotapes/month (one math, one science) from each teacher. If activities are longer than 30 minutes, the activity will be divided into two segments, to be separately coded. Coders will be randomly assigned to classrooms and to videotapes.

Analyses: Quality and Fidelity, Child Outcomes, and Change Over Time. Children’s gains in mathematics and their end-of-year scores for science will be plotted against both classroom quality and curricular fidelity. Correlational analyses will help identify relationships between quality and fidelity, and between teacher/classroom characteristics and quality and fidelity. We will also investigate changes in quality and fidelity across the year.

Findings/Results & Conclusions:

At this point in time, we have yet to collect end-of-year data from direct child assessments and teacher ratings of children’s skills. Reported here are the beginning-of-year data for these measures. Baseline data for children’s Mathematics and Vocabulary performance are included in Table 3. All data are presented in terms of the raw scores achieved.

Child Outcomes-Mathematics. On the TEMA-3 (a measure of number sense and operations), the average number of problem solved correctly was 7 (SD = 5.44). Based on the standardization data for this measure, for students at this approximate age (from 4 years, 0 months to 4 years, 2 months), a score of 7 translates to a math ability score of 89. (The math ability scores are scaled identically to IQ scores, with a mean of 100 and a standard deviation of 15.) A score of 89, though still within one standard deviation of the mean, is still well below the mean for children of that age.

For a measure of children’s skills in geometry and measurement, we drew a subset of items from the EMA-G and supplemented with items adapted or newly developed to assess all aspects of geometry and measurement indicated in our curricular goals. Norms are not available.
for the resulting measure; we therefore report the raw scores. In the area of geometry/measurement, the average score was 24 (out of 50, $SD = 6.88$).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEMA-3</td>
<td>7.09</td>
</tr>
<tr>
<td>Geometry/Measurement*</td>
<td>24.39</td>
</tr>
<tr>
<td>PPVT-4</td>
<td>67.16</td>
</tr>
</tbody>
</table>

* Includes items from the EMA-G

Child Outcomes – Receptive Vocabulary. The average number of pictures correctly identified by children on the PPVT-4 was 67 ($SD = 18.55$). This score translates to a scaled score of 101 for a child of 4 years, 0 months to 4 years, 1 month, indicating that our sample has an average receptive vocabulary for this age group. (This measure also employs IQ-style scaling, with a scaled mean of 100 and standard deviation of 15).

Initial Between-Group Differences in Math and Receptive Vocabulary. A one-way Analysis of Variance (ANOVA) test was conducted to determine if any differences existed between the treatment and control groups on the measures of math and language. On all measures, the treatment and control groups were statistically similar.

Teacher Ratings: Mathematics. Teachers were asked to complete a modified version of the Mathematics Academic Rating Scale (given in the ECLS-K) for each child in their classroom. Seven of the eight teachers returned the forms, with teachers evaluating beginning of year performance for 119 children, out of 136 total. In the ARS, teachers rate each child on 12 different mathematic abilities, with individual item ratings ranging from 1 (“not yet demonstrating”) to 5 (“proficient”). Sum scores for children ranged from 12 (the lowest rating possible) to 56 (the highest sum score is 60). On average, students received a sum score of 31.61 ($SD = 12.71$) points, indicating children’s skills were generally perceived to be between “Beginning” and “In Progress.”

Other data from year two are still being collected and analyzed, including our evaluative data to inform curricular revisions, dosage, fidelity, and teaching quality. These results, together with end-of-year data from direct child assessments and teacher ratings, will be reported in a future paper, as will results from a year three field trial of our curricula and teacher supports.
References


