Development of an IRT-based Direct Assessment of Preschool Science
Ximena Dominguez, Ariela C. Greenberg, Michelle F. Maier, Janna M. Fuccillo, & Daryl B. Greenfield

Introduction
Recent research has indicated that very young children are able to use scientific thinking as a model to guide learning in other domains (NRC, 2005; National Research Council and Institute of Medicine, 2000). As a result, a number of preschool curricula with science as the foundational element have been developed (e.g., French, 2004; Gelman & Brenneman, 2004). Measures of preschool science are needed to evaluate science-based curricula and interventions, but currently few are available.

The purpose of this project is to develop a reliable and valid Item Response Theory (IRT)-based direct assessment of preschool science using the guidelines for measurement development from the Standards for Educational and Psychological Testing (AERA, APA and NCME, 1999). The development of this measure is part of a larger preschool science project being conducted in collaboration with the Miami Museum of Science’s Center for Interactive Learning (CIL).

Sample Portions of Test Blueprint

<table>
<thead>
<tr>
<th>Category</th>
<th>Subcategory</th>
<th>Knowing</th>
<th>Observing</th>
<th>Describing</th>
<th>Comparing</th>
<th>Questioning</th>
<th>Predicting</th>
<th>Experimenting</th>
<th>Reflecting</th>
<th>Cooperating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Skills</td>
<td>Content Free</td>
<td>Sample Item 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life Sciences</td>
<td>Living Things</td>
<td>Sample Item 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earth &amp; Space Sciences</td>
<td>Weather</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical &amp; Energy</td>
<td>States &amp; Changes</td>
<td>Sample Item 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Test Development: Current Steps

Step ① – Create test item pool
This test blueprint is being used to create 150 dichotomous items. Most items tap a process skill (e.g., Observing) and a content sub-category (e.g., Living Things in the category of Life Sciences). Some items refer only to a process skill. Prompts are verbal and response formats include verbal, pointing, sorting, and sequencing.

Step ② – Create a test blueprint
Standards, curricula and assessments were reviewed in order to document current expectations for preschool science:
• 29 national and state Pre-K/Kindergarten science standards
• 10 early childhood general and/or science curricula
• 1 teacher rating scale of preschool children’s science knowledge and skills

This information was first divided into science inquiry skills and science content areas and then organized into categories and subcategories. To create the test blueprint, content areas were listed as content categories (rows) and process skills as processing levels (columns). See Sample Portions of Test Blueprint.

Step ③ – Conduct a content review of the items
Experts in the fields of science, measurement, and early childhood will be recruited to review and rate items on developmental appropriateness, content validity, and difficulty level.

Step ④ – Pilot test and conduct preliminary item analyses
Fall 2007 – Items will be pilot tested with 200 three- and four-year-old Head Start children. Classrooms will be randomly selected to include both teachers who have and have not had science training. A one-parameter logistic IRT model will be run to assess the difficulty of each item. Items will be revised and eliminated as necessary.

Spring 2008 – Revised items will be pilot tested with 500 children. A two-parameter logistic IRT model will be run to assess item difficulty and discrimination. Items will be revised and eliminated as necessary.

Step ⑤ – Conduct reliability and validity analyses
IRT analyses will provide item and test reliability, as well as internal validity through item difficulty and discrimination. The scores will be correlated with preschool measures of math and literacy for criterion-related validity.

Step ⑥ – Develop guidelines
Guidelines for administration, scoring, and interpretation of the test will be written. The technical manual will include reliability and validity results.