

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).

Test Development: Completed Steps

Step ① - Identify the purpose of the test and the construct to be measured

- Purpose of the direct assessment is to measure preschool children's science skills and knowledge
- Constructs to be measured are science inquiry skills and science content

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*.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C050052 and Grant DED-R305C050052 to the University of Miami. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education.

Sample Portions of Test Blueprint

Category	Subcategory	Knowing	Observing	Describing	Comparing	Questioning	Predicting	Experimenting	Reflecting	Cooperating
Process Skills	Content Free		Sample Item 1							
Life Sciences	Living Things				Sample Item 2					
	Working with Plants						Sample Item 3			
Earth & Space Sciences	Weather									
Physical & Energy Sciences	States & Changes					Sample Item 4				

Test Development: Current Steps

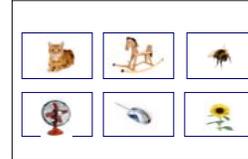
Step ③ – Create test item pool

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.



Sample Item 1

Assessor: "Your friend would like to see the stars at night. Point to what he can use to see them better."



Sample Item 2

Assessor: "Here are some pictures." (place cards according to the numbers on the back) "Give me all the pictures of things that are alive."



Sample Item 3

Assessor: "Sara watered the flower in her garden very well. What do you think will happen to the flower?"



Sample Item 4

Assessor: "Here are three pictures of water. Point to the picture of a solid."

Test Development: Future Steps

Step ④ – 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 ⑤ – 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 ⑥ - 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 ⑦ – Develop guidelines

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