An Assessment Primer: What is effective assessment in the early childhood classroom?

Dr. Shannon Riley–Ayers, National Institute for Early Education Research Rutgers University

Presented 2014 San Antonio Association for the Education of Young Children
NIEER conducts and communicates research to support high-quality, effective early childhood education for all young children.

The Institute offers independent research-based advice and technical assistance to policy makers, journalists, researchers, and educators.
Agenda

- Take a test!
- Review of Validity and Reliability
- Types of Assessments
- Performance-based Assessment Process
- Applying the Process
- Going Forward
Test Time!

- What is H₂O?
  - Water
- One bucket full of water contains more atoms than there are buckets full of water in the:
  - A. Bathtub
  - B. Swimming pool
  - C. Atlantic Ocean
- C. Atlantic Ocean
The human body contains enough carbon to provide enough “lead” (graphite) for about how many pencils?

9,000

Hot peppers get their heat from a molecule called capsacin. While the molecule acts as an irritant to mammals, including humans, _______ lack the receptor responsible for the effect and are immune to the burning sensation from exposure.

Birds
Score Interpretation

- 0–2 correct: in 5 years you will be a poor provider for young children
- 3–4 correct: in 5 years you will demonstrate characteristics of a ideal provider for young children
Wait a minute!
What were the problems?

- **Objectivity**
  - Scores are undistorted by biases of the individuals who administer and score them

- **Standard Conditions**
  - Administered and scored the same to increase objectivity
Validity Concerns

- Predictive Validity
  - The degree to which the predictions made by a test are confirmed by later behavior

- Concurrent Validity
  - The extent to which individual’s test scores on the new test correspond to scores on an established test of the same construct administered around the same time

- Construct Validity
  - The extent to which a particular test can be shown to assess the construct that it purports to measure

- Content Validity
  - The degree to which the scores yielded by a test adequately represents the content
Reliability Concerns

- Items only a sample of the total domain
- Failure to administer the test consistently
- Inconsistent in scoring procedures
- Unfavorable testing conditions
- Variability of how an individual feels
Does the assessment have a systematic approach and acceptable reliability and validity data? Has the assessment been used successfully with similar children?
What is Assessment?

Assessment is the process of gathering information about children from several forms of evidence, then organizing and interpreting that information.

- Evidence is an outward sign or indication of the child’s development or learning.

Why Do We Assess?

- Monitor children’s development and learning
- Guide interactions
- Make decisions about special services
- Report and communicate with others
### How do we Assess?

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Types of Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Formal</td>
</tr>
<tr>
<td>Preparedness to benefit from program</td>
<td>Readiness tests - criterion referenced</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify child who needs diagnostic test</td>
<td>Screening test with follow-up</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Determine sp. ed. classification</td>
<td>Diagnostic/developmental tests</td>
</tr>
<tr>
<td>To evaluate child progress</td>
<td>Standards-based achievement tests</td>
</tr>
</tbody>
</table>

Does the purpose of the assessment match the intended use of the assessment? Is the assessment appropriate for the age and background of the children to whom it will be administered?

Who administers the assessment?

Can we use Kindergarten assessments in PreK?
Why Use Observation-Based Performance Assessment?

- Learning is assessed in a natural environment.
- Skills and knowledge are captured in real life, over time (children are not “reliable”).
- Children are compared to themselves.
- Observation-based performance assessment is:
  - comprehensive,
  - focuses on strengths and interests,
  - is understandable to parents, and
  - informs teaching and learning.
Observer Issues

- Intra–observer reliability
  - The extent to which the observer is consistent in his or her observations and interpretations

- Inter–observer reliability
  - The extent that two observers agree with each other during actual data collection

- Criterion–related observer reliability
  - Extent to which a trained observer’s scores agree with those of an expert
POLICY QUESTION

- Are the data **easily collected** and interpreted to effectively inform teaching and learning?
- *How can teachers use assessments efficiently? (Teachers are using time at home to score.)*
Assessment occurs as an iterative cycle to help children reach standards.

Benchmarks to guide the assessment process.
POLICY QUESTIONS

- Are the necessary contextual **supports** in place for assessment implementation and effective, meaningful data use? (e.g., training, time, ongoing support)
- What are the policies for implementation and what is the roll-out **plan** for the assessment?
- Will data be gathered and maintained within FERPA and other **security** guidelines? Are there processes in place to inform stakeholders about how data are being gathered and held securely to allay concerns?
The Assessment Process: Where to Begin

Watch children in varied situations.

Be a participant observer — interact with the children.
Guidelines for Effective Observation

- Observe over time.
- Watch children in varied situations.
  - Social settings
  - Time of day
  - Individual preferences
  - Degree of choice
  - Level of competence
- Keep track of what you observe.
- Observe in and out of the action.

Record children’s behaviors with a focus on the identified items.

Collect work samples and record information.

Think about when and how to document.

Examine data for reflection.
Record what the child does without judgment or labeling (be sure to date the records).

Collect and date work samples and record information in an anecdote (reference the work sample by date at the end of the record).

Identify all the items on the assessment that each piece of evidence demonstrates.
When to Document

- In the moment with the child
- As soon as possible after the event
- Stepping out of the action in teamwork with your colleagues
- In reflection

Does the assessment allow the convergence of information from multiple sources/caregivers?

What technology is necessary in order to gather data?
Ty and Mark run right for the yellow ball. They both grab it and try to wrestle it from each other yelling so loud. They were both mad.
There is *not enough information* to be high-quality evidence.

The anecdote is not complete because it *does not provide the resolution* to the disagreement.

It also *does not provide objective descriptions* of children, but subjective (they were both mad).
Yalenis went to play. She went to the kitchen area. She wanted to play restaurant so she first picked up the menu in the bin. Then she went to the top of the refrigerator and got the paper and looked for a marker. She went to the writing table to get the marker. She went back to the kitchen area and picked up the menu and paper. She sat down with the menu and paper and marker and began to write. When she was done I asked her if I could have the paper. She said, “OK.” Then she put the menu back where she got it from and put the marker and the paper in the refrigerator. She left the area and went to the blocks.
Although this anecdote is long, it does not provide sufficient information. It is simply a running record of Yalenis’ movements for several minutes.

It does not provide information that would be meaningful to any items in the assessment.
The Assessment Process: Thinking about data

Examine data several times throughout the scoring period.

Continue observing and documenting.

Score and report results three times per year.
Analyzing and Evaluating: Improving Data

- Examine the data several times during the course of the score period to help focus your data collection.
  - Quantity and quality

- Two weeks before the end of the score period, examine the data and begin scoring.
  - This will highlight issues or holes in your data and will guide your observations for the last two weeks.
  - Do not stop observing or documenting at this time.
Evaluating: Scoring Procedures

- Scoring generally takes place three times a year.
- Organize and examine evidence.
- Place evidence on the continuum and provide a score if appropriate.
Does the assessment have a base or trajectory/continuum that is aligned to child developmental expectations, standards, and curricula? Does the assessment include all key domains?
The Assessment Process: Using Data

Form hypotheses based on the data.

Plan activities with differentiation based on data.

Communicate children's accomplishments and your plans for instructing with others.
Data Provided

- Qualitative data—use to improve instruction
- Quantitative data—use scores for reporting
- Vertical data—examine the child over time
- Horizontal data—examine the class data or center data
- Use data to communicate children’s accomplishments
Use the data collected and the results of your assessment to guide your interactions with the children.

Assessment and instruction are not separate acts, but rather work in concert.
POLICY QUESTION

- Are the data **useful** to teachers and other stakeholders?

- *How do we know what is the best assessment for our preschool children, we are driven by the state and adopted assessment. Do you have a tool we can use to review plans?*
1. Does the purpose of the assessment match the intended use of the assessment? Is the assessment appropriate for the age and background of the children to whom it will be administered?

2. Does the assessment allow the convergence of information from multiple sources/caregivers?

3. Are the necessary contextual supports in place for assessment implementation and effective, meaningful data use? (e.g., training, time, ongoing support)

4. Does the assessment have a base or trajectory/continuum that is aligned to child developmental expectations, standards, and curricula? Does the assessment include all key domains?

5. Does the assessment have a systematic approach and acceptable reliability and validity data? Has the assessment been used successfully with similar children?
6. Are the data easily collected and interpreted to effectively inform teaching and learning?
7. What technology is necessary in order to gather data?
8. Are the data useful to teachers and other stakeholders?
9. What are the policies for implementation and what is the roll-out plan for the assessment?
10. Will data be gathered and maintained within FERPA and other security guidelines? Are there processes in place to inform stakeholders about how data are being gathered and held securely to allay concerns?

# Geometry and Measurement

<table>
<thead>
<tr>
<th>Identifying and Using Shapes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifies circle and square</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takes apart and fits together objects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifies common shapes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turns and flips shapes intentionally to determine congruency or to solve a puzzle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifies additional shapes and irregular shapes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compares/contrasts 2- and/or 3-dimensional shapes by attribute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uses knowledge of shape properties to solve problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measurement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notices large differences in size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Makes direct comparisons of length, weight, volume, height, or area of materials/objects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uses standard and/or non-standard tools to measure length, height, volume, or weight</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uses a common base when comparing length or height</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypotheses</td>
<td>Activities</td>
<td>Interactions</td>
<td>Materials</td>
<td>Outcome</td>
<td></td>
</tr>
<tr>
<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>
</tr>
<tr>
<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>
</tr>
</tbody>
</table>
## Writing

<table>
<thead>
<tr>
<th>Writing Level</th>
<th>Composing</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>• May identify scribbling as “writing”&lt;br&gt;• Does not give meaning to writing</td>
<td>• Draws or scribbles</td>
</tr>
<tr>
<td>2</td>
<td>• Verbally labels own “writing” or drawing</td>
<td>• Makes forms that resemble letters&lt;br&gt;• May write his or her own name</td>
</tr>
<tr>
<td>3</td>
<td>• Provides dictation to an adult to be written on a piece of work</td>
<td>• Strings conventional letters together (other than his or her own name)</td>
</tr>
<tr>
<td>4</td>
<td>• Writes symbols for a purpose—to convey information or tell a story</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What’s next?

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Activities</th>
<th>Interactions</th>
<th>Materials</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FORMATIVE ASSESSMENT: GUIDANCE FOR EARLY CHILDHOOD POLICYMAKERS

By Shannon Riley–Ayers

WWW.CEELO.ORG

Early Learning Scale
NIEER
Lakeshore Materials
Contact Information

Shannon Riley–Ayers

sayers@nieer.org
848.932.4350