Empirical Studies of the Ounce and Work Sampling in Infancy and Early Childhood

Samuel J. Meisels
Erikson Institute
smeisels@erikson.edu

Co-Authors and Collaborators

WSHS
- Yange Xue, Mathematica Policy Research
- Melissa Shamblott, St. Paul Public Schools

Ounce
- Kristy Beachy-Quick, Erikson Institute
- Xiaoli Wen, Erikson Institute
- Ilona Helin, Erikson Institute
- Lisa Michael, Erikson Institute

The best way to evaluate a child’s performance is to study performance, not something else.

Curriculum-Embedded Assessments

Assessments based on information from children’s typical classroom performance.

The Work Sampling System

A continuous progress, instructional assessment that helps teachers document and assess children’s skills, knowledge, behaviors, and academic achievements from age 3 to Grade 6.

Preschool 4: Developmental Checklist

<table>
<thead>
<tr>
<th>Language and Literacy</th>
<th>L</th>
<th>W</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gains meaning by listening. (p. 5)</td>
<td>In Process</td>
<td>In Process</td>
<td>Proficient</td>
</tr>
<tr>
<td>2. Follows two- or three-step directions (p. 5)</td>
<td>In Process</td>
<td>In Process</td>
<td>Proficient</td>
</tr>
<tr>
<td>3. Demonstrates phonological awareness (p. 5)</td>
<td>In Process</td>
<td>In Process</td>
<td>Proficient</td>
</tr>
</tbody>
</table>

In this presentation Dr. Meisels will mention several assessments. He is an author of two of them—the Work Sampling System and the Ounce Scale—and a consultant to the company that publishes them, Pearson Education. These assessments will be cited solely to illustrate the points raised in the lecture, rather than to promote the assessments or imply an endorsement by any particular institution.
The best way to improve a child’s performance is to teach the child, not test the child.

Can Head Start and Pre-K Teachers Use Performance Assessments Accurately?

Purpose of Study
To evaluate the reliability and validity of Work Sampling for Head Start (WSHS) as a measure of children’s language, literacy, and mathematics.

Principal Study Questions
1. Is WSHS a reliable and valid means of evaluating children’s achievement and progress?
2. What does the WSHS Checklist add to predictions of children’s achievement over and above demographic variables?
3. Does WSHS identify children at risk for learning difficulties as accurately as normative measures?

Sample for the Study
- Pre-K: Head Start, school-, and community-based programs
- N = 112; Mean age = 4.6 years
- 16 classrooms (volunteer), 5 schools
- 62.5% African-American, 9% Hispanic; 95% low income; 54.5% male; 12% special needs

Preschool 4: Guidelines
The Work Sampling System

Language and Literacy
1. Listening
3. Phonological awareness
   Phonological awareness refers to the ability to hear and discriminate the sounds of language. From now on, children are asked to find and distinguish the similar units of sound within words with teacher support. They are then asked to hear and discriminate rhymes, the beginning sounds of words, and rhyming words. This activity helps children develop phonemic awareness.
   - reading the room the teacher says and finding a word to rhyme with it
   - hearing the sound of the first letter in their own names and using this ability to sound out or “read” unknown “names” that begin with the same letter
   - understanding rhymes, giving them new rhyming words
   - counting the number of rhymes in their names
   - hearing the sound of a letter when they use it in a new word because they recognize it from a familiar word (for example, the “s” sound in “sing”)
   - hearing, rhyming words, and creating rhyming words and nonsense words, such as “tune, tune, money, money, money, tune”
   - counting familiar rhymes with classmates during circle time or at play.
Outcomes Measures

- Test of Early Reading Ability (3d edition)
- Test of Early Mathematics Ability (3d edition)

Results of the WSHS Validity Study (Pre-K)

1. WSHS has high internal reliability (alphas = .90 – .94) and moderately high correlations between fall and spring ratings (.71 for literacy and .65 for math).

Results of the WSHS Validity Study (Pre-K), cont.

2. WSHS correlates moderately well (.30 - .44) with standardized, individually administered psychoeducational tests of early reading and math (TERA and TEMA).

Mean Scores on Normative Tests (S.D.)

<table>
<thead>
<tr>
<th>Test</th>
<th>Fall</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEMA-3</td>
<td>82.79 (13.42)</td>
<td>83.77 (14.41)</td>
</tr>
<tr>
<td>TERA-3</td>
<td>90.27 (11.43)</td>
<td>90.67 (11.78)</td>
</tr>
</tbody>
</table>

Results of the WSHS Validity Study (Pre-K), cont.

3. After adjusting for demographics, WSHS adds unique information (~20%) to predictions about children’s early reading and math achievement.

Results of the WSS Validity Study (Pre-K), cont.

4. Data obtained from WSS accurately discriminate between children who are and are not at risk.
Conclusions - 1

WSHS has excellent internal consistency and reliability within domains across time.

Conclusions - 2

WSHS is a valid and effective assessment of children’s learning.

Conclusions - 3

WSHS ratings accurately identify children at risk for learning difficulties in literacy and math.

Functional Assessment

Assessments that focus on everyday, naturally occurring, practical behaviors and accomplishments

The Ounce Scale

Basic Characteristics

• A relationship-building tool
• Enhances parent and provider knowledge, control, and empowerment
• Helps differentiate and expand parents’ and providers’ perceptions of babies

Observation Record

Toddlers I: 18–24 months

1. Personal Connections: It’s about trust

How does the toddler show that they are engaged, involved, and interested in the activity?

The toddler might do one or more of the following:

- Fans themselves, such as with hair, dress, or necktie
- Points at objects in the environment
- Brysure
- Puts water on the face, puts it on the head, or puts it on the foot
- Points at objects in the environment
- Puts hands over eyes or covers their mouth
- Shakes a toy or puts it in their mouth

How does this behavior show that the toddler is engaged in the activity?

Describe the actions the toddler is taking to show engagement in the activity.
Developmental Profile
Toddlers I: 18–24 months

How accurately do Early Head Start teachers use observational assessments?

The Ounce Scale
Research Questions

1. Is the Ounce a reliable and valid assessment of learning and development?
2. Does the Ounce identify children in need of special instruction or services?
3. Does the Ounce have an impact on teaching and caregiving?

Sample for the Validation Project

- Birth – 42 months, cross-sectional
- N = 251 children in 8 age groups
- 124 classrooms (volunteer) from 7 Early Head Start programs
- 63% African-American, 26% Hispanic; 92% low income; 55% male; 11% special needs (IFSP)

Outcome Assessments

- Bayley Scales of Infant Development-II Mental and Motor Scale;
- Preschool Language Scale-4;
- ASQ:Social Emotional (completed by teachers)

Results of the Study

1. Reliability: Moderate (overall Cronbach alpha = .65 [range = .19-.89])
2. Validity: Correlations between Ounce Developmental Profiles and criterion measures of mental, motor, and language are moderate (.28 - .32) but higher for social-emotional (.47).
Correlations between Ounce Scales and Criterion Measures

<table>
<thead>
<tr>
<th>Ounce Scale</th>
<th>ASQ:SE</th>
<th>PLS-4</th>
<th>Bayley-II Mental</th>
<th>Bayley-II Motor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social-Emotional</td>
<td>.47**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td>.32**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognition</td>
<td></td>
<td>.31**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motor</td>
<td></td>
<td></td>
<td>.28**</td>
<td></td>
</tr>
</tbody>
</table>

*p < .05, **p < .01

Results of the Study - 2

3. The more seriously delayed the child, as rated by the Ounce, the less competently the child performed on the outcomes.

4. The Ounce significantly predicted performance over and above demographics, especially on social-emotional.

Does the Ounce Distinguish Between Children At-Risk and Not At-Risk?

More than 70% of the infants and toddlers who are at risk, as defined by the standardized assessments, were correctly identified.

“Needs Development” Rating

Teachers appeared to systematically avoid the rating, Needs Development, particularly for younger children in the sample.

Teacher Interview Study: Description of Sample

- 21 teachers from five EHS centers (73% response rate) + 7 program managers
- All completed at least two Observation Records and Developmental Profiles that were used in the Ounce Scale Validation Study
- Range of teachers’ education: highest = Masters degree, lowest = high school
- 16 African American, 3 Hispanic, 9 White (all female)

Teacher Interview Study: Content of Questions

1. How teachers use the different components and ratings of the Ounce Scale
2. How teachers use the Ounce Scale for planning, setting goals, and working with parents
3. Training and support teachers were given in use of the Ounce Scale
4. Global questions, e.g., has the Ounce Scale changed them as teachers, do they like it?
Reasons For Not Using “Needs Development”

1. Teachers viewed this rating negatively
2. Teachers had difficulty applying the standards to atypical behavior
3. The age range may have made caregivers reluctant to say that a child needs development
4. No intermediate rating levels were available
5. Teachers wanted to protect parents from “bad news”

How Did the Ounce Affect Teachers and Caregivers?

1. Assessing children
2. Knowledge and understanding of child development
3. Setting goals
4. Communication with parents
5. Relationships with co-workers

Teacher Characteristics That Are Associated with Greater Accuracy in Use of the Ounce

• Higher levels of education
• More training in early care and education
• Being relatively new to working with infants and toddlers
• Longer use of the Ounce Scale

Understanding the child within context is key to understanding the child