Evidence Base of Commonly Used Early Childhood Assessments

Teresa Brown, M.Ed. & Sophia Hubbell, M.A.T.
Kent State University
Brooke Winchell, M.Ed., ABD
University of Wisconsin – Whitewater
Presented at DEC Annual Conference, Kansas City, MO
October 16, 2010

Session Overview

- What are Assessment Standards
- What do we mean by evidence base
- How we evaluated evidence
- Results of our research

Session Goals

- To learn how to be savvy consumers of assessments
- To understand the current state of evidence available for some commonly-used assessments
Assessment Standards/Guidelines

- NAEYC (National Association for the Education of Young Children)
- DEC (Division for Early Childhood)
- Head Start
- State Standards
- AERA (American Education Research Association)
- LINK (Neisworth, Bagnato, & Pretti-Frontczak, 2010)

DEC, NAEYC, & Head Start Guidelines

- Assessment should be:
  - Comprehensive
  - Authentic
  - Conducted for ethical and meaningful purposes
  - Systematic well planned and evidence based
  - Involve families
  - Conducted in a developmentally appropriate manner
  - Reflect the child’s language and culture
AERA Standards

Include 24 Standards related to validity

- Standard 1.1
  A rationale should be presented for each recommended interpretation and use of test scores together with a comprehensive summary of the evidence and theory bearing on the intended use or interpretation.

- Standard 1.3
  If validity for some common or likely interpretation has not been investigated or if the interpretation is inconsistent with available evidence, that fact should be made clear and potential users should be cautioned about making unsupported interpretations.

(AERA, APA, & NCME, 1999)

AERA Standards

- Standard 1.4
  If a test is used in a way that has not been validated, it is incumbent on the user to justify the new use collecting new evidence if necessary.

- Standard 1.13
  When validity evidence includes statistical analyses of test results either alone or together with data on other variables the conditions under which the data were collected should be described in enough detail that users can justify the relevance of the statistical findings to local conditions. …

- Standard 1.22
  When it is clearly stated or implied that a recommended test use will result in a specific outcome the basis for expecting that outcome should be presented together with relative evidence.
LINK Standards

- Acceptability
- Authenticity
- Collaboration
- Evidence
- Multi-factors
- Sensitivity
- Universality
- Utility

(Neisworth, Bagnato, & Pretti-Frontczak, 2010)

Evidence Base Components

- Professional Standards
- Diversity Representation
- Disability Specific
- Early Intervention Validation

(Neisworth, Bagnato, & Pretti-Frontczak, 2010)
Selection of Assessments

Top 35 based on LINK consumer ratings

Removed:
- Single Domain Assessments
- Infant/Toddler Only Assessments
- Disability Only Assessments

LINK final 9

- New Portage
- H/S COR
- Creative
- AEPS
- BSSI-3
- Galileo
- DOCS
- HELP
- TPBA-2

Selection of Assessments

Top 9
Linking Authentic Assessment and Early Childhood Intervention: Best Measures for Best Practices (Bagnato, Neisworth, & Pretti-Frontczak, 2010)

Top 10
Part C Child Outcomes Measurement System Activities Table (The Early Childhood Outcomes Center, August, 2010)

Final 10
Early Childhood Assessment in an Age of Accountability (Pretti-Frontczak & Brewer, 2005)

8 Keeping It “R-E-A-L” with Authentic Assessment (Macy & Bagnato, 2010)
What Counts as Evidence

- **Study** - a manuscript that describes the methods and results (according to APA) of an empirical investigation of technical adequacy, utility, or fidelity, or related to an assessment.
  - Indexed online
  - Easily accessible to the general public.
  - Published in a journal, component of the published assessment (e.g. manual), on a product-related website, or otherwise available in full-text form online, including theses and dissertations
  - Investigations of Technical Adequacy include all types of reliability, validity, and bias.
Common Reasons for Exclusion

- Subjects not in target age group (i.e. preschool)
- Inadequate or non-existent methods description
- Curriculum

Validation for Intended Purposes

<table>
<thead>
<tr>
<th></th>
<th>Eligibility</th>
<th>Programming</th>
<th>Outcomes Evaluation</th>
<th>Accountability</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEPS</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Battelle</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Brigance</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Creative</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DOCS</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Galileo</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>HELP</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>H/S COR</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>TPBA-2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>WSS</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
### Evidence Rating Scale

<table>
<thead>
<tr>
<th>Rating</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Considerable evidence exists to validate the tool for the intended purpose.</td>
</tr>
<tr>
<td>Medium</td>
<td>Some evidence exists to validate the tool for the intended purpose.</td>
</tr>
<tr>
<td>Low</td>
<td>Little evidence exists to validate the tool for the intended purpose.</td>
</tr>
<tr>
<td>None</td>
<td>No evidence exists to validate the tool for the intended purpose.</td>
</tr>
</tbody>
</table>

### Results

<table>
<thead>
<tr>
<th></th>
<th>Eligibility</th>
<th>Programming</th>
<th>Outcomes Evaluation</th>
<th>Accountability</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEPS</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Battelle</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>N/A</td>
</tr>
<tr>
<td>Brigance</td>
<td>N/A</td>
<td>Low</td>
<td>Low</td>
<td>None</td>
</tr>
<tr>
<td>Creative</td>
<td>N/A</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>DOCS</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>N/A</td>
</tr>
<tr>
<td>Galileo</td>
<td>N/A</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>HELP</td>
<td>N/A</td>
<td>None</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>H/SCOR</td>
<td>N/A</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>TPBA-2</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>N/A</td>
</tr>
<tr>
<td>WSS</td>
<td>N/A</td>
<td>None</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
Quantity ≠ Quality

- We did not examine the psychometric properties of the given studies.
- Sample size and characteristics (age, disability, etc…) should be taken into consideration. At least some of the research should include samples with characteristics that match those of the children you intend to assess with the measure.
- Future research should look at quality and outcomes of the studies.

References


References


Contact Information

Please feel free to contact us for questions or more information.

- Teresa Brown, trummer@kent.edu
- Sophie Hubbell, shubbell@kent.edu
- Brooke Winchell, winchelb@uww.edu

- Handouts will be available online through the DEC conference website.