BEST PRACTICE IN THE ASSESSMENT OF ENGLISH LEARNERS

Monica Oganes, M.A. Ed.S.
Licensed School Psychologist

Alabama Council of Administrators of Special Education Conference
Montgomery, Alabama
February 23, 2016

Learning Targets

Participants will be able to:
- assess English Learner’s educational and linguistic needs
- select appropriate curriculum-based and standardized assessment tools to evaluate English Learners
- interpret assessment results considering linguistic and cultural diversity factors


Copyright Monica Oganes
Alabama’s English Learners

<table>
<thead>
<tr>
<th>2012-13 Data</th>
<th>USA</th>
<th>Alabama</th>
</tr>
</thead>
<tbody>
<tr>
<td>EL students</td>
<td>9.6% of total enrollment</td>
<td>2.6% of total enrollment</td>
</tr>
<tr>
<td>Top 5 Home Languages</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 1. Spanish | 66.8% | Spanish | 7.7%
| 2. Chinese | 6.3% | Arabic | 2.9%
| 3. Vietnamese | 3.0% | Korean | 2.6%
| 4. Korean | 2.5% | Chinese | 1.9%
| 5. Tagalog | 1.9% | Vietnamese | 1.9%

Home Languages (% of total EL):
- Spanish: 65.8%
- Chinese: 6.1%
- Vietnamese: 3.3%
- Korean: 2.5%
- Tagalog: 1.9%

Enrollment by Race (% of all enrolled):
- Asian: 4.4%
- Black: 14.7%
- Hispanic: 22.1%
- American Indian: 1.0%
- Pacific Islander: 0.2%
- White: 68.5%

Free/Reduced Lunch (% of all enrolled):
- Free: 43.0%
- Reduced: 6.6%

Students with Disabilities (% of all enrolled):
- Free: 14.6%
- Reduced: 7.0%


BICS and CALP (Cummins, 1981)

**BICS** (1-2 years)
- Basic Interpersonal Communication Skills
  - Conversational Fluency
  - Pronunciation
  - Vocabulary
  - Grammar

**CALP** (5-7 years)
- Cognitive Academic Language Proficiency
  - Abstract language for academic work
  - Complex linguistic ability
  - Analysis, Synthesis, Evaluation

First and Second Language Acquisition

BICS

CALP

Common Underlying Proficiency

Cummins (1981)

State Education Resource Center (SERC) 2009
Factors That Threaten Assessment Validity

- The most pervasive bias in testing is not related to test items, test structure, test reliability, factor structure, or predictive validity
- The greatest threat is incorrect interpretation, which introduces faulty evaluative judgment to tests that may not be biased
- Acculturative knowledge: a diverse child's cultural background experiences are different from those in the normative sample
- Developmental language proficiency: a linguistically diverse child's level of second language acquisition and academic language proficiency
- The use of assessment procedures that decrease the validity of the derived score

As presented by H. Ochoa and S. Ortiz at the 2015 NASP Annual Conference, "Current Best Practices in Assessment and Intervention with English Learners" (2/20/2015).

Faulty Evaluative Judgment

- Labeling a child as disabled without ruling out exclusionary factors (Faulty logic)
- Insisting that a child is responding to intervention even though the trend line shows insufficient or poor response (Error of attack)
- Recommending the use of a particular program based on the publishers' claims of success when independent research does not back the claims (Error of weak reference)
- Insisting that a low IQ score on a test administered in English to a non English speaker is a valid measure of cognitive ability (Error of misinformation)
- Assuming a statistically significant difference is rare without checking base rates (Statistical limitations)

Marzano and Brown (2009)

Consultation

- Multicultural consultation enhances teachers' knowledge, skills, and confidence in working with diverse families
- Incorporates language and cultural considerations at the individual, group, family, and systems levels
- Consultation allows sharing of knowledge
- Consultees plan, adapt, and deliver effective services

Adapted from NASP Position Statement for the Provision of School Psychological Services to Bilingual Students, 2015
Construct Validity:
an assessment's ability to measure a
construct that it was designed to
assess or that it purports to measure
Llorente (2008)

Reliability:
• Consistency or degree of accuracy of evaluation
  results
• Impacted by each of its components (variables)
• Overall reliability (Rx) is the multiplicative product
  of the reliability of each of its components
• As good as the lowest reliability coefficient of any
  of its components

Cognitive Equivalency = Construct
Validity

Digit Span: Test of auditory simple attention

• Numbers words do not have the same number of syllables
• Differential impact on memorization and later recall: 5-8-1-4
• Performance differences amongst ethnic groups unrelated to
differences in cognitive abilities

5 = five  
5 = cin-co  
One syllable  
Two syllables

Examples of Reliability and Validity

Variables

Oral language proficiency (background
knowledge)
Use of culturally loaded English assessments
with English Learners (OWLS)
Vocabulary
Use of untrained interpreters
Inadequate curriculum used at Tiers I, II, III
Lack of cognitive equivalency (Digit Span)
Variance (e.g., CALP vs. BICS, ethnicity/culture)
Assessment

- Use of valid and reliable assessment tools and procedures
- Use of native language
- Nonverbal does not guarantee validity/reliability
- Norms for native language tests may not represent ELLs in U.S. schools, limited research
- Training in the use of interpreters and awareness of reliance on interpreters.

Adapted from NASP Position Statement for the Provision of School Psychological Services to Bilingual Students, 2015

Woodcock-Johnson-IV Oral Language Battery

- Standardized battery includes multiple subtests to assess Oral Language Development in English, including receptive and expressive language, phonemic awareness, and retrieval fluency.
- It includes three equated Spanish subtests to derive a comparable Oral Language score.
- The battery provides Cognitive Academic Language Proficiency scores as well as Relative Proficiency Indexes
- It can be used in conjunction with the Achievement and Cognitive batteries to assess other broad and narrow abilities and skills.

WJ-IV Achievement Battery and BATERIA-III Achievement

- The WJ-IV assesses all areas of reading, writing and math, providing CALP scores and Relative Proficiency indexes.
- It is co-normed with the WJ-IV Cognitive and Oral Language batteries, allowing for the assessment of broad and narrow abilities, and clusters of abilities and skills based CHC.
- The BATERIA-III also assesses the same academic areas, but it is not equated to the WJ-IV. Has subtests similar to the WJ-III but there are subtests that have been changed; while others have been eliminated or added. Caution must be used when comparing Spanish (BATERIA-III) and English (WJ-IV) scores, as they are different versions, they are not equated, and the interpretation criteria for RPI has been changed. BATERIA-IV is in the works.
Test Selection

Determine Student's Oral Language Proficiency

- WIDA: Become familiar with language development standards
- Woodcock Johnson Oral Language Battery
  - English and Spanish CALP can be obtained
  - Dominant language determines what tests to select
  - Verbal subtests most affected by not having fluent language proficiency in English
  - Performance subtests may have instructions in English—difficult to understand task for a non-proficient child. Do not assume that student can do performance tasks when English tests are used.
  - Non-Spanish speakers can be given English proficiency test to determine fluency level and make decisions.
  - If oral language proficiency is low in native language and English, consultation with a speech/language pathologist is recommended.

How does Alabama measure English Proficiency?

- WIDA ACCESS Placement Test (W-APT™), with WIDA ACCESS for ELL's
  - W-APT™ assesses proficiency in all four language domains as well as comprehension
  - Kindergarten children with a score of 25 or above are considered proficient, but their scores need to be monitored to see if a re-screening is needed in 1st grade.
  - Overall Composite scores of 4.0 and below identify the child as Limited English Proficient and requires placement in an English language instruction educational program
  - Overall Composite scores of 4.0 and above may identify the child as LEP and further evaluation is warranted.
  - W-APT™ Composite scores are used to match the child to the level on the WIDA ACCESS for ELLs, which measures performance on a continuous proficiency scale.
  - WIDA ACCESS for ELLs tests measure grade-level cluster and range of proficiencies in relation to the proficiency continuum, allowing for measurement of growth over time.

- WIDA MODEL Proficiency Assessments (optional use at district expense)
  - New English Language assessments for K-12 aligned to the Language Proficiency Standards
  - Assess ability in four language domains and evaluate Social and Instructional English and Academic Language for different areas
  - Re-aligned by curricular and instructional guidance, CAN DO descriptors, and performance indicators
  - Can be used for identification (screener when first enrolled), progress monitoring (benchmark assessments during the year), or summative assessment (given annually for reporting purposes)

Test Selection

- Nonverbal tools or native language tests do not guarantee valid and reliable data.
- Nonverbal tests rely on some form of effective communication between examiner and examinee, and may be as culturally loaded as verbal tests—limited validity of results.
- Native language tests may not represent the types of ELLs typically found in the U.S.—limited research.

Adapted from NASP Position Statement for the Provision of School Psychological Services to Bilingual Students, 2015
Test Selection

- Examples of Nonverbal Tests:
  - Universal Nonverbal Intelligence Test, 2nd Edition (UNIT2)
  - Wechsler Nonverbal Scale of Ability (WNV)
  - Comprehensive Test of Nonverbal Intelligence, 2nd Ed. (CTONI-2)
  - Leiter-3
  - Test of Nonverbal Intelligence, 4th Edition (TONI-4)

- Examples of Language-Reduced Tests (instructions in English)
  - Differential Ability Scales Special Nonverbal Composite
  - Performance Scales

What other tests do you administer?

Test Selection

Academics:
- If a child has had recent formal instruction in native language, use a measure in native language if possible.
- If academic testing in native language not available, obtain history of academic performance in home country.
- If instruction is provided in English only in the U.S., academic measures in English can be given to obtain baseline data.
- Math is not universal. Other countries teach math differently (i.e., centimeters vs. inches).

Test Selection

- Cognitive process test selection will depend on student’s language proficiency.
- Language-based subtests (i.e., List memory) cannot be given in English if child dominant in native language.
- Language-based subtests cannot be translated into native language (Digit Span is language-based also).
- Test Instructions in English also interfere with a child’s ability to understand a task.
- Do not translate tests on the spot.

The greatest threat in validity is incorrect interpretation, which introduces faulty evaluative judgment to tests that may not be biased.
Case Study Review

Take a few minutes to review the case provided on your own:
- Circle critical information and data to be considered.
- Make a list of exclusionary factors to be investigated.
- Underline strengths and protective factors.

For reproduction, please contact Monica Oganes
www.monicaoganes.com
monica@monicaoganes.com