Development of an Assessment of High Frequency English Vocabulary for Young English Language Learners

Lauren Artzi, Lindsey Massoud, and Diane August, Center for Applied Linguistics; Chris Barr, University of Houston

Goal

- Develop an item type to assess young English learners’ knowledge of different kinds of words. The young English learners (ELs) were in second grade classrooms in U.S. schools.

Background

- Identified three issues with extant vocabulary measures:
  1. Individually administered → cumbersome testing scenario
  2. Lack of nonlinguistic support → harder for lower level ELs
  3. Images appear without explanation → difficult to test abstract words

- Group administered to second grade Spanish-speaking English learners (n=184 in year 1, n=187 in year 2) in a transitional bilingual education program in a large urban district in the Southwest
- Administered as part of a two-phased vocabulary intervention study as a pretest and posttest curriculum-based, researcher-developed measure
- Assessed student knowledge of words that are frequent in grade-level text (according to The Educator’s Word Frequency Guide [Zeno, et al., 1995]) but with above-grade-level meanings, that aren’t well known until sixth grade (according to the Living Word Vocabulary [Dale & O’Rourke, 1981])

Test Development

- Developed four sub-tests with a total of 72 items in year 1 and 84 items in year 2, stratified on attributes associated with acquisition: Spanish-English cognate status and conceptual complexity.

Methods

- Conceptual Complexity
  - Less complex: More difficult to perceive with the senses
  - More complex: Less imageable
  - More imageable: High related
  - Non-related: Requires more knowledge of other words/concepts

- Types of Words Tested
<table>
<thead>
<tr>
<th>Cognate</th>
<th>Non-Cognate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptually simple</td>
<td>18 items (year 1) 21 items (year 2)</td>
</tr>
<tr>
<td>Examples: applied delicate quantity singular</td>
<td>Examples: clung illness opposite weary</td>
</tr>
<tr>
<td>Conceptually complex</td>
<td>18 items (year 1) 21 items (year 2)</td>
</tr>
<tr>
<td>Examples: confidence responsibility informed preferred</td>
<td>Examples: actually dreadful judgment quality pride proper</td>
</tr>
</tbody>
</table>

Example Items

Example Item 1: Pride

Paragraph Read Aloud by the Test Administrator

Put your finger on number 15. Number 15. Inge got a 100 on her test, so she feels good about her work.

When someone feels good about something they did, do they say they feel “rage,” “concern,” or “pride”?

Listen again and bubble in the word that means when someone feels good about something they did: “rage,” “concern,” or “pride.”

Student Answer Sheet

15

feels good about something they did

rage o concern o pride

Example Item 2: Dreadful

Paragraph Read Aloud by the Test Administrator

Put your finger on number 1. Number 1. These players feel very bad because they lost their game.

When someone feels very bad, do they say they feel “dreadful,” “remarkable,” or “enthusiastic”?

Listen again and bubble in the word that means when someone feels very bad: “dreadful,” “remarkable,” or “enthusiastic.”

Student Answer Sheet

1

dreadful o remarkable o 

Validity Evidence

- Found significant correlations between our researcher-developed assessment and established measures:
  - Year 1: Gates-MacGinitie Word Knowledge test
    - Pre-test r=.49, p <.01; Post-test r=.60, p <.01
  - Year 2: TOLD Oral Vocabulary subtest
    - Pre-test r=.49, p <.01; Post-test r=.66, p <.01

- Would expect higher correlations if word meanings being tested were on grade level rather than above grade level

Discussion

- The advantages of this approach are the following:
  1. It is multimodal—children hear a story with a definition, see a picture and a definition.
  2. Unlike assessments that use just pictures, this approach allows us to assess a variety of word types, including abstract, less imageable vocabulary.
  3. The assessment can be group administered at lower grade levels because of the scaffolding provided.

Reliability

- Computed coefficient alphas to investigate reliability of the assessment:
  - Year 1: Pre-test, α = .63; Post-test, α = .88
  - Year 2: Pre-test, α = .78; Post-test, α = .91

- Would expect reliability to look similar to post-test reliability if words tested were on grade level rather than above grade level

Future Directions

- Develop a measure using this item type for grade-level words

Acknowledgements

University of Houston: Colin Carlson, Ken Nisser, Xio Gutierrez

Center for Applied Linguistics: Domy Kenyon, Inge Sigellner, Cheryl Dressler, Aiken Bach, Annie Duguay, Kat Kramer

Harvard University: Paula Uccelli, Catherine Snow

This project effort was supported by Grant Number 01HD03950 from the National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Child Health and Development or the National Institutes of Health.