Linguistic Modification of Assessments for ELL Students
U.S. Department of Education, LEP Partnership

Jamal Abedi

University of California, Davis

National Center for Research on Evaluation, Standards, & Student Testing, UCLA Graduate School of Education

October 28, 2007

Are there Assessment Issues Specific to English Language Learners?

- Studies have shown that academic achievement tests that are constructed for native English speakers have lower reliability and validity for ELL students (Abedi, Leon, & Mirocha, 2003).
- Linguistic and cultural biases may negatively impact validity of assessments for these students.
- Therefore, results of these tests may not be interpreted for ELL students the same way as they are for non- ELL students.
- These assessment issues may seriously affect classification, curriculum, instruction and accommodation decisions for these students.

Why Assessment is So Important for ELL Students?

For ELL students assessment starts before instruction

Assessment results affect ELL students in the following areas:

- *Classification*
- *Instruction*
- *Accountability (the NCLB issues)*
- *Promotion*
- *Graduation*

Thus assessment of ELL students is very high stakes.

How do ELL students do in assessments in comparison with non-ELL students?

- ELL students perform lower than non-ELL students in general
- The performance-gap between ELL and non-ELL students increases as the language demand of test items increases
- The performance-gap approaches zero in content areas with a minimal level of linguistic complexity (e.g. math computation)

ARE THE STANDARDIZED ACHIEVEMENT TESTS APPROPRIATE FOR ELLS?

The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) elaborated on this issue:

For all test takers, any test that employs language is, in part, a measure of their language skills. This is of particular concern for test takers whose first language is not the language of the test. Test use with individuals who have not sufficiently acquired the language of the test may introduce construct-irrelevant components to the testing process. (p. 91)

Are the Standardized Achievement Tests Reliable and Valid for these Students?

- The reliability coefficients of the test scores for ELL students are substantially lower than those for non-ELL students
- ELL students' test outcomes show lower criterion-related validity
- Structural relationships between test components and across measurement domains are lower for ELL students

Site 2 Stanford 9 Sub-scale Reliabilities (Alpha), Grade 9

Sub-scale (Items)	English Only	LEP
Reading, N=	181,202	52,720
-Vocabulary (30)	.835	.666
-Reading Comp (54)	.916	.833
Average Reliability	.876	.750
Math, N=	183,262	54,815
-Total (48)	.898	.802
Language, N=	180,743	52,863
-Mechanics (24)	.803	.686
-Expression (24)	.812	.680
Average Reliability	.813	.683
Science, N=	144,821	40,255
-Total (40)	.805	.597
Social Science, N=	181,078	53,925
-Total (40)	.805	.530

Why these tests are not reliable for ELL students

There must be additional sources of measurement error affecting the assessment outcome for these students

These sources include:

- Linguistic complexity of test items
- Cultural factors
- Interaction between linguistic and cultural factors with other student background variables

Language of Assessment

- A clear and concise language is a requirement for reliable and valid assessments for ELL students
- It may also be important consideration for students with learning disabilities since a large majority of students with disabilities are in the *Learning Disability* category
- Students in the *Learning Disability* category may have difficulty processing complex language in assessment
- Simplifying the language of test items will also help students with disabilities, particularly those with learning disabilities

Examining Complex Linguistic Features in Content-Based Test Items

Feature	Feature Description	Categories Combined
1	Item length	1, 2, 4, 45
2	Vocabulary	3, 26, 27
3	Nominal heaviness	5, 6, 29, 30, 31, 32
4	Verb voice	7, 33
5	Modal	8, 34
6	Relative clause	9, 10, 11, 35, 36, 37
7	Adverbial modification	12, 13, 14, 15, 16, 17, 38, 39, 40, 41
8	Conditional clause	18, 19
9	Complement clause	20, 44
10	Sentence structure	28, 42, 43, 46
11	Preferred argument structure	22, 23, 47, 48
12	Question form	21
13	Global difficulty	24
14	Content interest	25

Linguistic Modification Concerns

- **Familiarity / frequency of non-math vocabulary**: unfamiliar or infrequent words changed
  census > video game
  A certain reference file > Mack's company

- **Length of nominals:** long nominals shortened
  last year's class vice president > vice president
  the pattern of puppy's weight gain > the pattern above

- **Question phrases**: complex question phrases changed to simple question words
  At which of the following times > When
  which is best approximation of the number > approximately how many

Linguistic Modification cont.

- **Voice of verb phrase:** passive verb forms changed to active
  The weights of 3 objects were compared > Sandra compared the weights of 3 rabbits
  If a marble is taken from the bag > if you take a marble from the bag

- **Conditional clauses:** conditionals either replaced with separate sentences or order of conditional and main clause changed
  If Lee delivers x newspapers > Lee delivers x newspapers
  If two batteries in the sample were found to be dead > he found three broken pencils in the sample

- **Relative clauses:** relative clauses either removed or re-cast
  A report that contains 64 sheets of paper > He needs 64 sheets of paper for each report

Original Item

Harriet, Jim, Roberto, Maria, and Willie are in the same eighth grade class. One of them is this year's class president. Based on the following information, who is the class president?

The class president was last year's vice president and lives on Vince Street.
Willie is this year's class vice president.
Jim and Maria live on Cypress Street.
Roberto was not last year's vice president.

A. Jim
B. Harriet
C. Roberto
D. Maria
E. Willie

Modified Item

Harriet, Jim, Roberto, Maria, and Willie ran for president of their eight-grade class. One of them won. Who is president?

The president now was vice president last year and lives on Vince Street.
Willie is vice president now.
Jim and Maria live on Cypress Street.
Roberto was not vice president last year.

A. Jim
B. Harriet
C. Roberto
D. Maria
E. Willie

Example

Original:

A certain reference file contains approximately six billion facts. About how many millions is that?

A. 6,000,000
B. 600,000
C. 60,000
D. 6,000
E. 600

Modified:

Mack's company sold six billion pencils. About how many millions is that?

A. 6,000,000
B. 600,000
C. 60,000
D. 6,000
E. 600

Example

Original:

The census showed that three hundred fifty-six thousand, ninety-seven people lived in Middletown. Written as a number, that is:

A. 350,697
B. 356,097
C. 356,907
D. 356,970

Modified:

Janet played a video game. Her score was three hundred fifty-six thousand, ninety-seven. Written as number, that is:

A. 350,697
B. 356,097
C. 356,907
D. 356,970

- CRESST Studies on the Assessment and Accommodation of ELL Students:

Impact of Language Factors On Assessment of ELLs A Chain of Events

Fourteen studies on the assessment and 3 on the instruction (OTL)of ELL students

Study #1

Analyses of extant data (Abedi, Lord, & Plummer, 1995).

Used existing data from NAEP 1992 assessments in math and science.

SAMPLE: ELL and non-ELLs in grades 4, 8, and 12 main assessment. NAEP test items were grouped into long and short and linguistically complex/less complex items.

Findings

ELL students performed significantly lower on the longer test items.

- ELL students had higher proportions of omitted and/or not-reached items.
- ELL students had higher scores on the linguistically less-complex items.

Study #2

Interview study (Abedi, Lord, & Plummer, 1997)

37 students asked to express their preference between the original NAEP items and the linguistically modified version of these same items. Math test items were modified to reduce the level of linguistic complexity.

Findings

- Over 80% interviewed preferred the linguistically modified items over the original version.

Many students indicated that the language in the revised item was easier:

- “Well, it makes more sense.”
- “It explains better.”
- “Because that one’s more confusing.”
- “It seems simpler. You get a clear idea of what they want you to do.”

The revised items need less time for response:

- "It's easier to read, and it gets to the point, so you won't have to waste time."
- "I might have a faster time completing that one 'cause there's less reading."
- "Less reading; then I might be able to get to the other one in time to finish both of them."
- "'Cause it's, like, a little bit less writing."

Study #3

Impact of linguistic factors on students' performance (Abedi, Lord, & Plummer, 1997).

Two studies: testing performance and speed.

SAMPLE: 1,031 grade 8 ELL and non-ELL students.
41 classes from 21 southern California schools.

Findings

- ELL students who received a linguistically modified version of the math test items performed significantly better than those receiving the original test items.

Study #4

The impact of different types of accommodations on students with limited English proficiency (Abedi, Lord, & Hofstetter, 1997)

SAMPLE: **1,394** grade 8 students. 56 classes from 27 California schools.

Findings

Spanish translation of NAEP math test.

- **Spanish-speakers taking the Spanish translation version performed significantly lower than Spanish-speakers taking the English version.**
- **We believe that this is due to the impact of language of instruction on assessment.**

Linguistic Modification

- **Contributed to improved performance on 49% of the items.**

Extra Time

- **Helped grade 8 ELL students on NAEP math tests.**
- **Also aided non-ELL students. Limited potential as an assessment accommodation.**

Study #5

Impact of selected background variables on students' NAEP math performance (Abedi, Hofstetter, & Lord, 1998).

SAMPLE: **946** grade 8 ELL and non-ELL students. 38 classes from 19 southern California schools.

Findings

- Four different accommodations used (linguistically modified, a glossary only, extra time only, and a glossary plus extra time).
- The glossary plus extra time was the most effective accommodation.

Glossary plus extra time accommodation

- Non-ELLs showed a greater improvement (16%) than the ELLs (13%).
- This is the opposite of what is expected and casts doubt on the validity of this accommodation.

Study #6

The effects of accommodations on the assessment of LEP students in NAEP (Abedi, Lord, Kim, & Miyoshi, 2000)

SAMPLE: **422** grade 8 ELL and non-ELL students. 17 science classes from 9 southern California schools. A customized dictionary was used.

Findings

- Included only non-content words in the test.
- Customized dictionary easier to use than published dictionary.
- ELL students showed significant improvement in performance.
- No impact on the non-ELL performance.

Study #7

Language accommodation for large-scale assessment in science
(Abedi, Courtney, Leon, Mirocha, & Goldberg, 2001).
SAMPLE: **612** grades 4 and 8 students. 25 classes from 14 southern California schools.

Findings

- A published dictionary was both ineffective and administratively difficult as an accommodation.
- Different bilingual dictionaries had different entries, different content, and different format.

Study #8

Language accommodation for large-scale assessment in science
(Abedi, Courtney, & Leon, 2001)

SAMPLE: **1,856** grade 4 and **1,512** grade 8 ELL and non-ELL students.
132 classes from 40 school sites in four cities, three states.

Findings

- Results suggested: linguistic modification of test items improved performance of ELLs in grade 8.
- No change on the performance of non-ELLs with modified test.
- The validity of assessment was not compromised by the provision of an accommodation.

Study #9

Impact of students' language background on content-based performance: analyses of extant data (Abedi & Leon, 1999).

Analyses were performed on extant data, such as Stanford 9 and ITBS
SAMPLE: Over **900,000** students from four different sites nationwide.

Study #10

Examining ELL and non-ELL student performance differences and their relationship to background factors (Abedi, Leon, & Mirocha, 2001).

Data were analyzed for the language impact on assessment and accommodations of ELL students.

SAMPLE: Over **700,000** students from four different sites nationwide.

Findings

- The higher the level of language demand of the test items, the higher the performance gap between ELL and non-ELL students.
- Large performance gap between ELL and non-ELL students on reading, science and math problem solving (about 15 NCE score points).
- This performance gap was zero in math computation.

Some of our Recent publications summarizing findings of our research on the assessment of ELLs :

- *Abedi, J. and Gandara, P. (2007). Performance of English Language Learners as a Subgroup in Large-Scale Assessment: Interaction of Research and Policy. Educational Measurement: Issues and Practices. Vol. 26, Issue 5, pp. 36-46.*
- *Abedi, J. (2007). Utilizing accommodations in the assessment of English language learners. In: Encyclopedia of Language and Education. Heidelberg, Germany: Springer Science+ Business Media.*
- *Abedi, J. (2007). English Language Learners with Disabilities. In Cahlan, C. & Cook, L. Accommodating student with disabilities on state assessments: What works? (Ed.) New Jersey: Educational Testing Service.*
- *Abedi, J. (2006). Psychometric Issues in the ELL Assessment and Special Education Eligibility. Teacher's College Record, Vol. 108, No. 11, 2282-2303.*
- *Abedi, J. (2006). Language Issues in Item-Development. In Downing, S. M. and Haladyna, T. M. Handbook of Test Development (Ed.). New Jersey: Lawrence Erlbaum Associates, Publishers.*
- *Abedi, J. (2005). Assessment: Issue and Consequences for English Language Learners. In Herman, J. L. and Haertel, E. H. Uses and Misuses of Data in Accountability Testing (Ed.) Massachusetts: Blackwell Publishing Malden.*

Conclusions and Recommendation

Assessment for ELL students:

- Must be based on a sound psychometric principles
- Must be controlled for all sources of nuisance or confounding variables
- Must be free of unnecessary linguistic complexities
- Must include sufficient number of ELLs in its development process (field testing, standard setting, etc.)
- Must be free of biases, such as cultural biases
- Must be sensitive to students' linguistics and cultural needs