

# *Academic Assessment of English Language Learners: Foundational Research*

*John W. Young, Ph.D.  
Senior Research Scientist  
Center for Validity Research  
Educational Testing Service*

# *What is the main goal of our research efforts?*

The long-range goal is to ensure that all content area assessments designed for use in grades K – 12 (Title 1 assessments) are valid and fair for all examinees, including English Language Learner (ELL) examinees.

# ***What is needed to achieve this goal?***

There are three sequential steps being followed:

- Develop a conceptual foundational research framework
- Create a research agenda for ELL examinees based on this framework
- Conduct the necessary empirical studies on assessments taken by ELL examinees

# *A Research Framework for Validity, Quality, and Fairness*

A foundational research framework should investigate the following aspects of any assessment:

- ***Candidates and Populations:*** Understand who takes the test
- ***Psychometric Properties:*** Examine psychometric characteristics at the item and test level

# *A Research Framework for Validity, Quality, and Fairness*

- ***Score Interpretation:*** Evaluate score inferences and their use
- ***Validity:*** Gather evidence to evaluate the intended inferences and actions based on reported scores
- ***Fairness and Accessibility:*** Evaluate technical quality and fairness for all examinees, including ELLs

# *A Research Framework for Validity, Quality, and Fairness*

- ***Scores and Scales:*** Evaluate existing score scales (including equating issues) and the development of new scales
- ***Policy Issues:*** Evaluate the uses of test scores for decision-making as they impact individuals and institutions

# ***A Research Agenda for ELL Examinees***

“In evaluating the validity of a test for a subgroup, the primary concern is test fairness, and the measure of fairness is comparability (Willingham et al., 1988).”

Adopting this perspective, the main emphasis of a research agenda for ELL examinees should be on fairness and comparability.

# *A Research Agenda for ELL Examinees*

In evaluating assessments taken by ELL examinees, there are 8 indicators of comparability:

1. ***Reliability***: Equal precision of measurement
2. ***Factor Structure***: Relationships among test components are similar across examinee groups

# *A Research Agenda for ELL Examinees*

3. ***Differential Item Functioning (DIF)***: No differential item functioning due to group membership
4. ***Predictive Validity***: No differential prediction due to group membership
5. ***Educational Decisions***: No differential decision-making due to group membership

# *A Research Agenda for ELL Examinees*

6. ***Test Content:*** Content and cognitive processes are similar across examinee groups
7. ***Testing Accommodations:***  
Accommodations are appropriately used, perceived as such, and have the desired effects
8. ***Test Timing:*** No differential speededness due to group membership

# *The Role of ETS*

ETS must ensure that all of our assessments meet the highest standards for technical quality in terms of validity, fairness, and accessibility for all examinees, including ELL examinees.

In addition, ETS strives to provide global leadership on important assessment and measurement issues.



# ***ETS Contributions to the Research Agenda for ELL Examinees***

***Candidates and Populations:*** Examining demographics of ELL examinees who take ETS-developed assessments in order to create a framework for understanding and interpreting these demographics. This framework will enable us to examine the influence that various demographic variables have on test performance.



# *ETS Contributions*

***Psychometric Properties:*** Analyses of ETS-developed state standards-based assessments in English language arts, mathematics, and science in grades 4, 5 and 8 identified few items (< 2%) exhibiting DIF against ELL examinees.

# *ETS Contributions*

***Psychometric Properties:*** A study of a standards-based mathematics assessment for 4<sup>th</sup> graders found that items with greater non-mathematical lexical and syntactic complexity had greater differences in IRT difficulty estimates favoring non-ELLs over ELLs. However, the impact of linguistic complexity is reduced for items with non-linguistic *schematic* representations.

# *ETS Contributions*

***Score Interpretation:*** Conducting a study using data from several standards-based mathematics tests to investigate the sensitivity of DIF methods to the linguistic complexity of the scores used to match examinees. DIF analyses using a linguistically simplified matching criterion variable may be more effective in identifying items exhibiting DIF against ELLs.

# *ETS Contributions*

***Validity:*** Exploratory and confirmatory factor analyses of ETS-developed state standards-based assessments for 5<sup>th</sup> and 8<sup>th</sup> graders in mathematics and science identified all assessments as being unidimensional for native English speakers and ELLs. However, there appears to be more construct-irrelevant noise in the first factor, especially for ELL examinees without accommodations.

# *ETS Contributions*

***Validity:*** This study of ETS-developed state standards-based assessments also found that two testing accommodations for ELLs (translated directions; access to glossaries) had desirable effects on the dimensionality of the assessments. This provides evidence that these accommodations are beneficial to ELL examinees.

# *ETS Contributions*

***Validity:*** Conducting a validity study of the performance of two former ELL groups (students classified as initially fluent in English or reclassified as English fluent) on ETS-developed state standards-based assessments. To date, little research has been conducted on these two categories of students.

# *ETS Contributions*

***Validity:*** Seeking to partner with one or more school districts with large numbers of ELLs to carry out a longitudinal study of ELL and non-ELL students with regard to grades, retention, placement and classification decisions. The study would address equity questions concerning educational decisions using test scores and other indicators.

# *ETS Contributions*

***Fairness and Accessibility:*** Completing a study of linguistic modification (LM) as an ELL testing accommodation. LM principles were identified and applied by ETS test developers to disclosed mathematics and science items. Original and modified versions of test items have been field tested with samples of ELLs and non-ELLs in grades 4 and 6 in three states.



# *ETS Contributions*

***Scores and Scales:*** Conducting a study to investigate the impact of linguistic complexity of test items on score equating for ELLs for several mathematics tests. The purpose of this study is to determine whether differences occur in the equating functions using subsets of linguistically simple items and linguistically complex items.

# *ETS Contributions*

***Policy Issues:*** Producing a set of guidelines that specifies best practices in the assessment of ELLs, particularly for those in grades K – 12. This document will be available in 2008.



# *In Summary*

The creation of a comprehensive long-range agenda guides us in carrying out a systematic program of research on the validity and fairness of academic assessments for all examinees.

Implementation of the research agenda will ensure that academic assessments will be of the highest technical quality and are valid and fair for all examinees, including ELLs.

# *Questions? Comments?*

## *Contact Information:*

*[jwyong@ets.org](mailto:jwyong@ets.org)*

*(609) 734-5037*

