

Explaining Differential Item Functioning on Academic Vocabulary Assessments for English Language Learners using Explanatory Item Response Models

Paulina Kulesz, Research Assistant Professor

University of Houston



Closing the Gaps for All Learners

Annual Principal Investigators Meeting

January 8-9, 2020 | Washington, DC

Problems with Assessing Academic Achievement among English Language Learners


- The percentage of English Language Learners in public schools in the United States is growing rapidly.
- Accurate and reliable assessments of English Language Learners academic achievement is a national concern.
- The poor psychometric performance of current assessments may be attributed to:
 - Language characteristic that differentially affect English Language Learners academic performance,
 - Many standardized assessments are normed for English-only population.

Current Sample

		Language Proficiency Status			
		EO (n = 3600)	IFEP (n = 1034)	RFEP (n = 3793)	ELL (n = 1851)
Year 1 2010-2011	6 th	982	220	908	532
	7 th	967	275	1029	518
	8 th	921	322	1051	469
	All	2870	817	2988	1519
Year 2 2011-2012	6 th	614	194	713	243
	7 th	70	13	48	52
	8 th	46	10	44	37
	All	730	217	805	332

Word Generation Academic Vocabulary Test

- Synonym task
- 50 items on each of two forms (years 2010-2011 and 2011-2012), some unique (81 items) and some shared across forms (19 items)
- Four response options (key and three distractors)
- Distractors: semantically, phonologically, or orthographically related, or unrelated



The diagram shows a vertical stack of four colored boxes within a rounded rectangular frame. From top to bottom: an orange box labeled 'TARGET', a green box labeled 'Key' with a yellow star to its right, a light gray box labeled 'Distractor A', and another light gray box labeled 'Distractor B'. A fourth light gray box labeled 'Distractor C' is positioned below the others but is not enclosed in the frame.

XX. He acquired a pet.

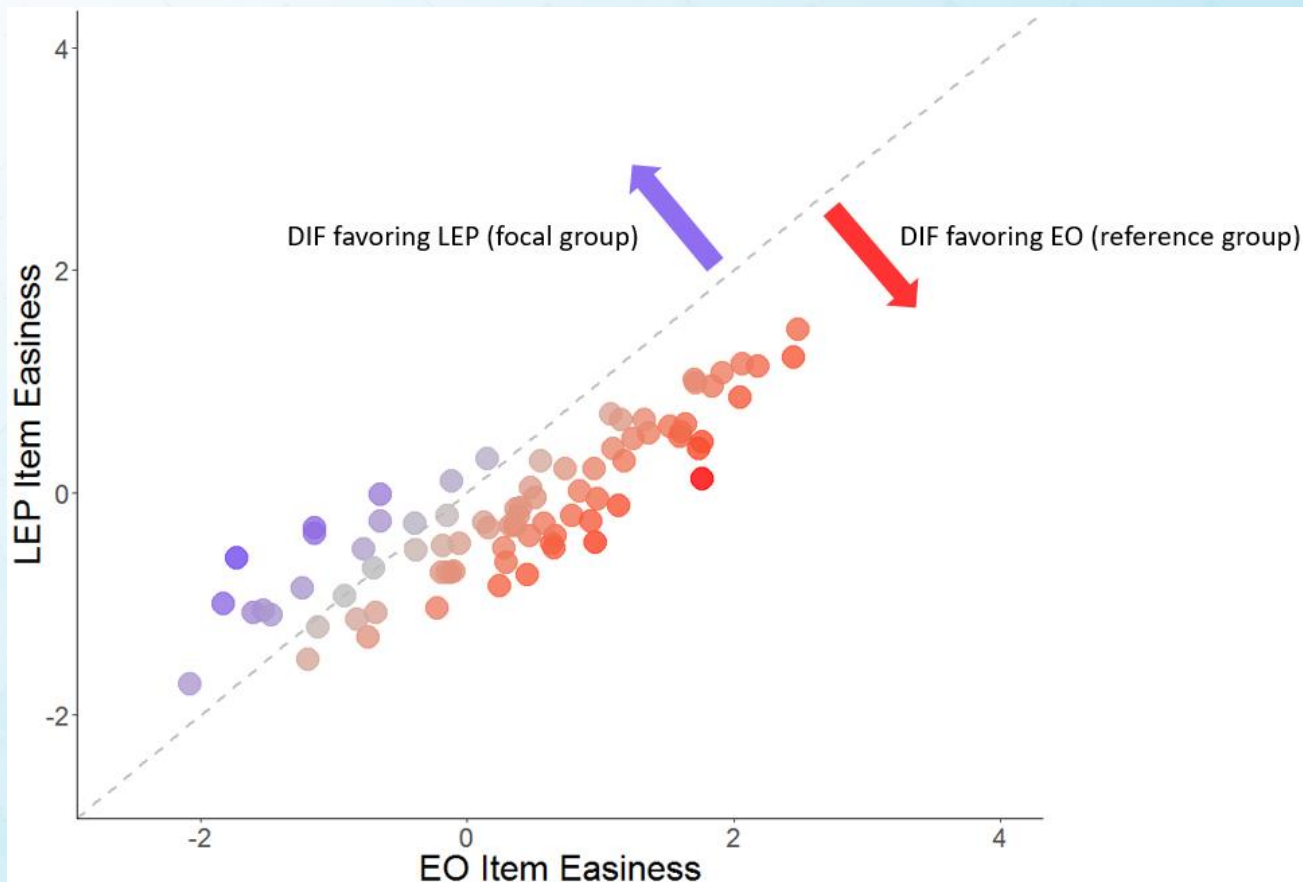
- a. got
- b. trained
- c. lost
- d. adored

Differential Item Functioning Analyses

- In the current context one-parameter models (uniform differential item functioning) were used as they:
 - Provide greater sensitivity,
 - Allow for analysis of items on both test forms simultaneously using a common items equating approach to link the two forms due to the presence of 19 overlapping items.
- Comparing groups: do two students with the same underlying ability have equal probabilities of getting an item correct?
 - Reference group: English-only
 - Focal group: Limited English Proficiency, Initially Fluent, or Reclassified

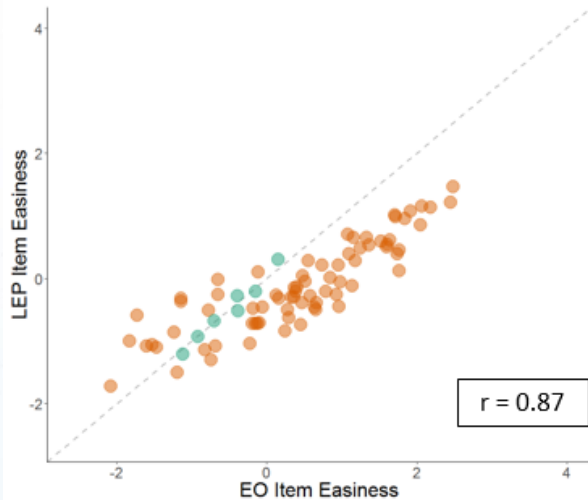
Interpreting Differential Item Functioning Results

- Current analyses modeled predicted item “easiness” (higher values = easier items) for each group

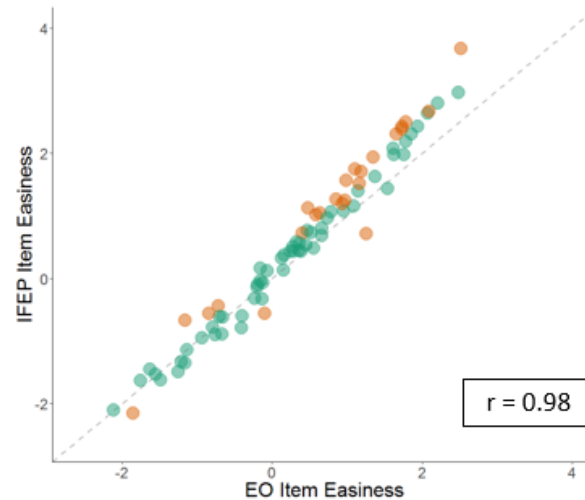


Differential Item Functioning Results

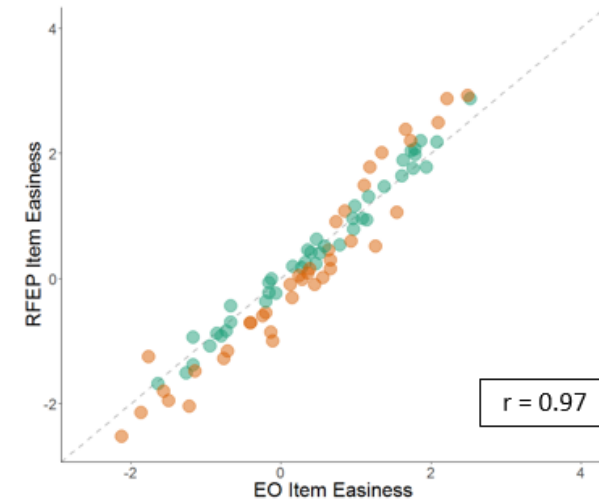
EO vs. LEP



EO vs. IFEP



EO vs. RFEP

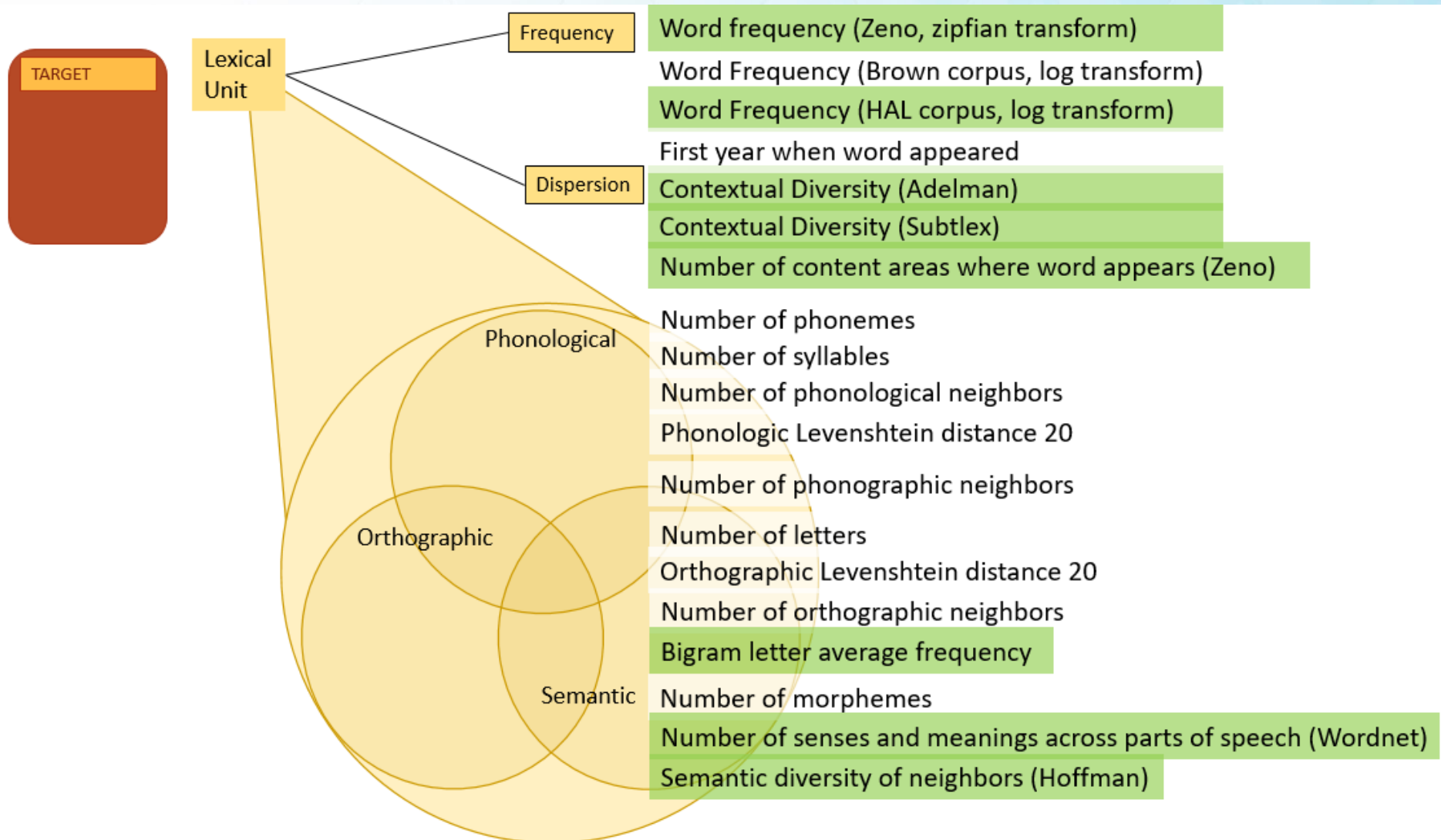


Primary focus of subsequent analyses was on English-only versus Limited English Proficiency contrast, since Initially Fluent and Reclassified groups showed difficulty estimates that were highly correlated with English-only estimates

Explanatory Item Response Models, Part 1

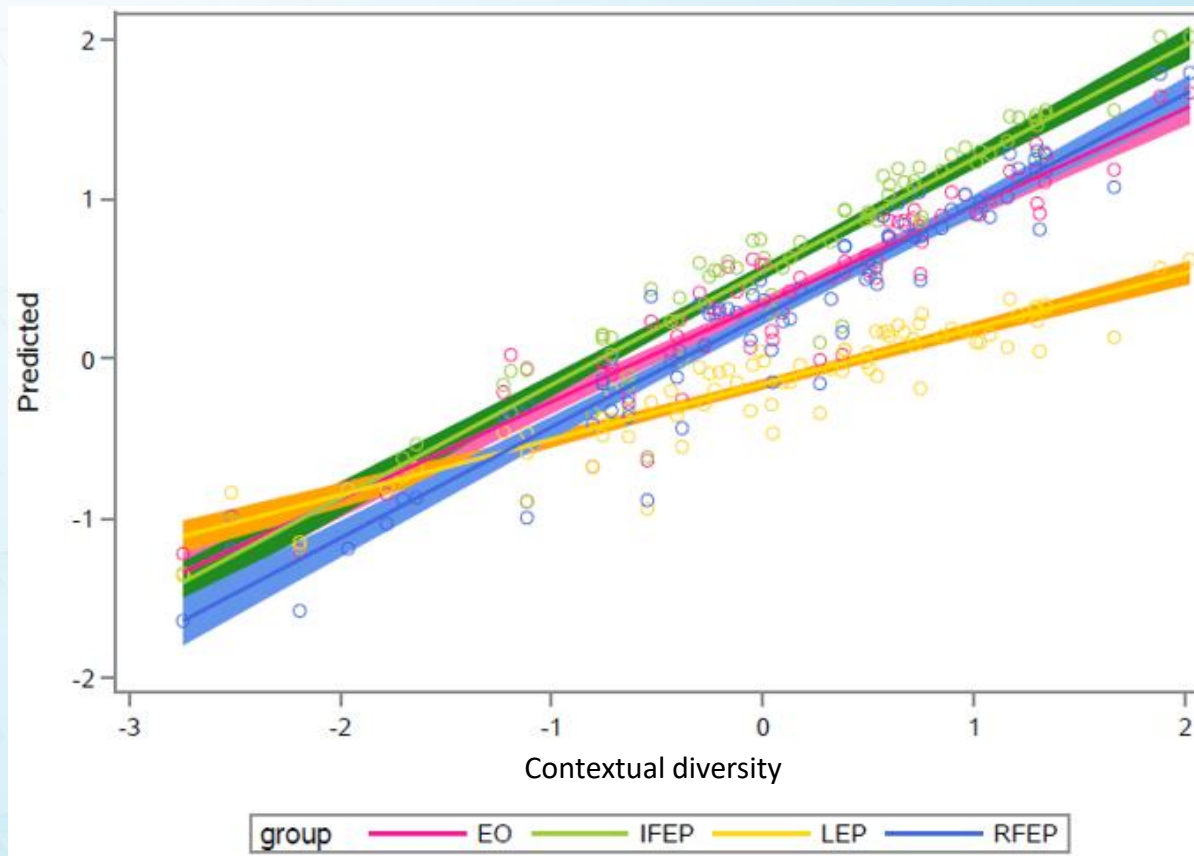
- Two-stage data analytic approach:
 - Saving values of predicted item easiness from differential item functioning analyses
 - Explaining predicted item easiness using target word characteristics and student characteristics (language proficiency status) as predictors
- Focusing on target word characteristics by language proficiency status interactions

Explanatory Item Response Models, Part 2



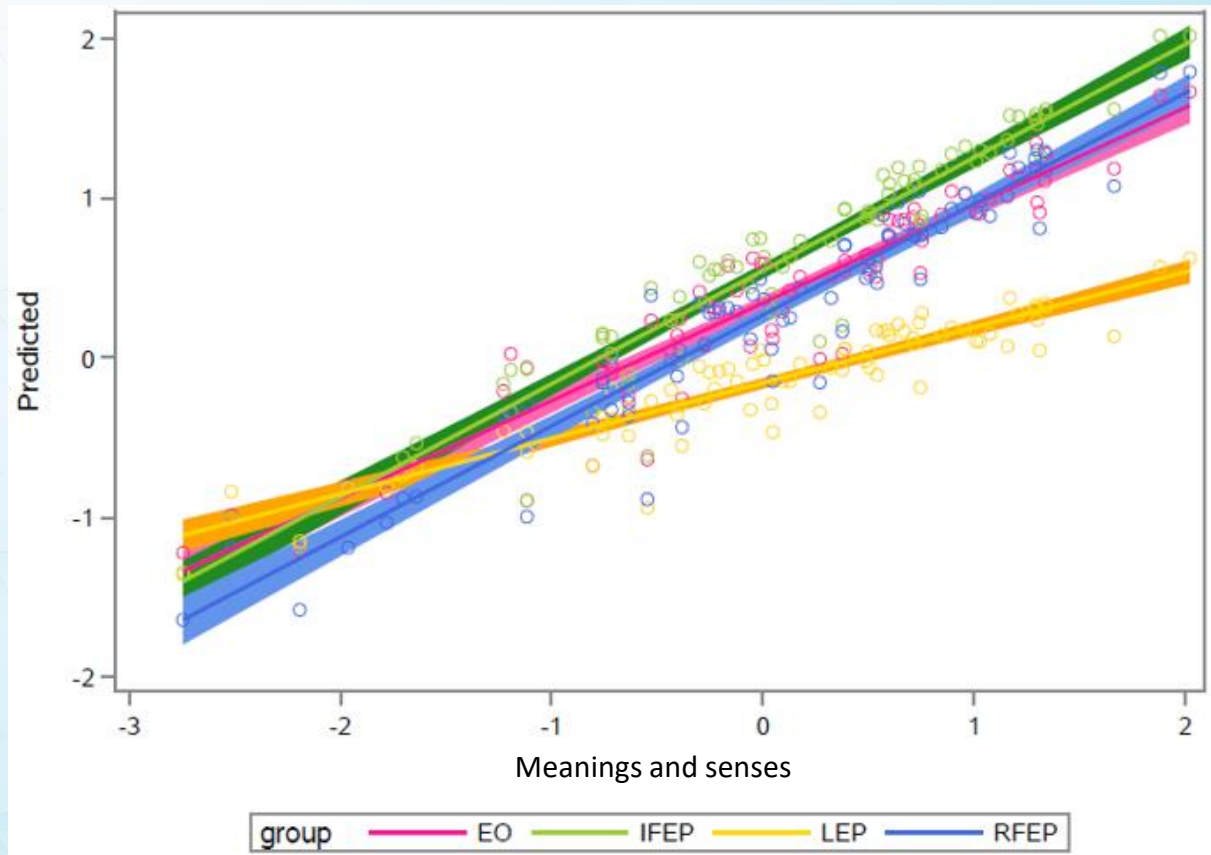
Results of Explanatory Item Response Models, Part 1

Contextual diversity by language proficiency status/group interaction



Results of Explanatory Item Response Models, Part 2

Meanings and senses by language proficiency status/group interaction



Discussion of Findings, Part 1

- Contextual diversity (subtitles): number of documents in the corpus containing a given word (corpus of subtitles from movies and television)
 - Single strongest predictor of item easiness for all groups.
 - Positive relation: greater contextual diversity is associated with easier items
 - Correlated with some other predictors (e.g., frequency), often the only significant unique predictor in models with multiple predictors

Discussion of Findings, Part 2

- Word frequency: higher frequency associated with easier items
- Number of meanings and senses (WordNet): combined meanings and senses across parts of speech
 - Greater number of meanings and senses associated with easier items

Conclusions

- Contextual diversity versus Frequency: where we encounter a word versus how often
- More meanings & senses: easier items
- How many predictors are enough?
- Initially Fluent & Reclassified students: show similar pattern to English-only students

Acknowledgement

The research reported here was supported by the Institute of Education Sciences, US Department of Education, through grants R305A170151 and R305A090555. The opinions expressed are those of the authors and do not represent the views of the Institute or the US Department of Education.