

# Differential Performance on Academic Vocabulary Assessments for English Language Learners (ELLs)

---

*Paulina A. Kulesz, Autumn L. McIlraith, Joshua Lawrence, &  
David J. Francis*

*AERA Conference*

*April 7, 2019*

 **TIMES**

# Overview

---

- Problems with Assessing Academic Achievement among ELLs
- Sample
- Differential Item Functioning (DIF) Analyses
- Relations between DIF and Item Characteristics

# Problems with Assessing Academic Achievement among ELLs

---

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

# Sample ( $N = 10,278$ )

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

*Note.* EO = English-only; IFEP = Initially fluent English proficient; RFEP = Redesignated fluent English proficient; LEP = Limited English Proficiency.

# Differential Item Functioning (DIF) Analyses

---

- The question of differential item functioning was addressed using:
  - *Logistic Regression (LR) Approach*
  - *Item Response Theory (IRT) Approach*

# Logistic Regression (LR) Approach

---

- The LR analyses were separately computed for each Academic Vocabulary (WG) item within test forms for Year 1 and Year 2.
  
- The LR analyses were also split into three sets depending on the type of total test score being used as a vocabulary knowledge ability estimate:
  - WG test score,
  - Gates test score,
  - Latent score.

# Logistic Regression (LR) Approach

---

**Uniform DIF** if Model 2 provided a better fit relative to Model 1

**Model 1:**

$$Y = \beta_0 + \beta_1 \text{Vocabulary Knowledge Ability}$$

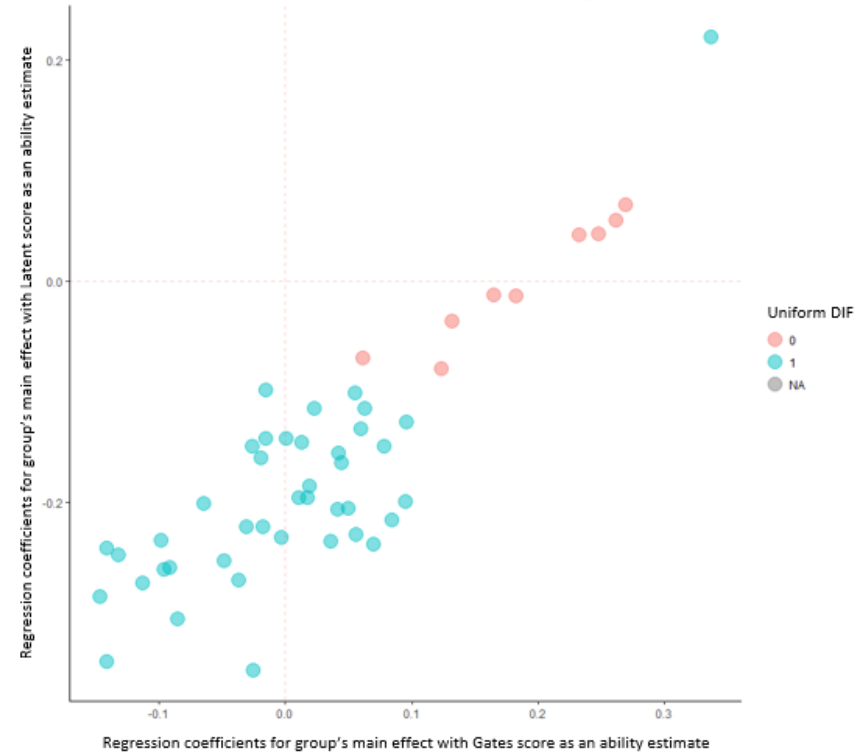
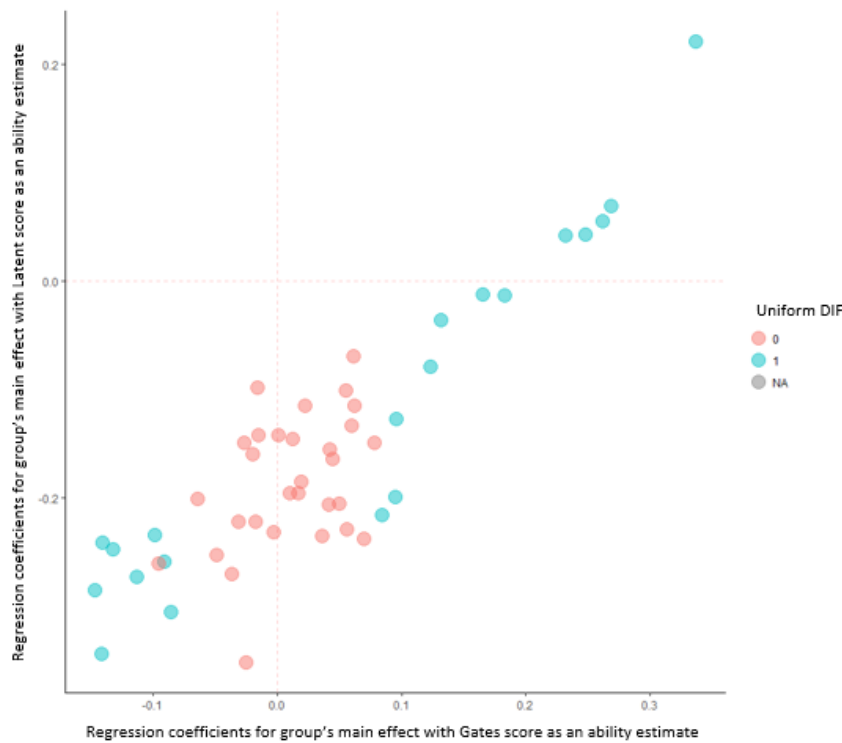
**Model 2:**

$$Y = \beta_0 + \beta_1 \text{Vocabulary Knowledge Ability} + \beta_2 \text{Group}$$

# Uniform DIF Results (LR Approach)

Gates score as an ability estimate

Latent score as an ability estimate



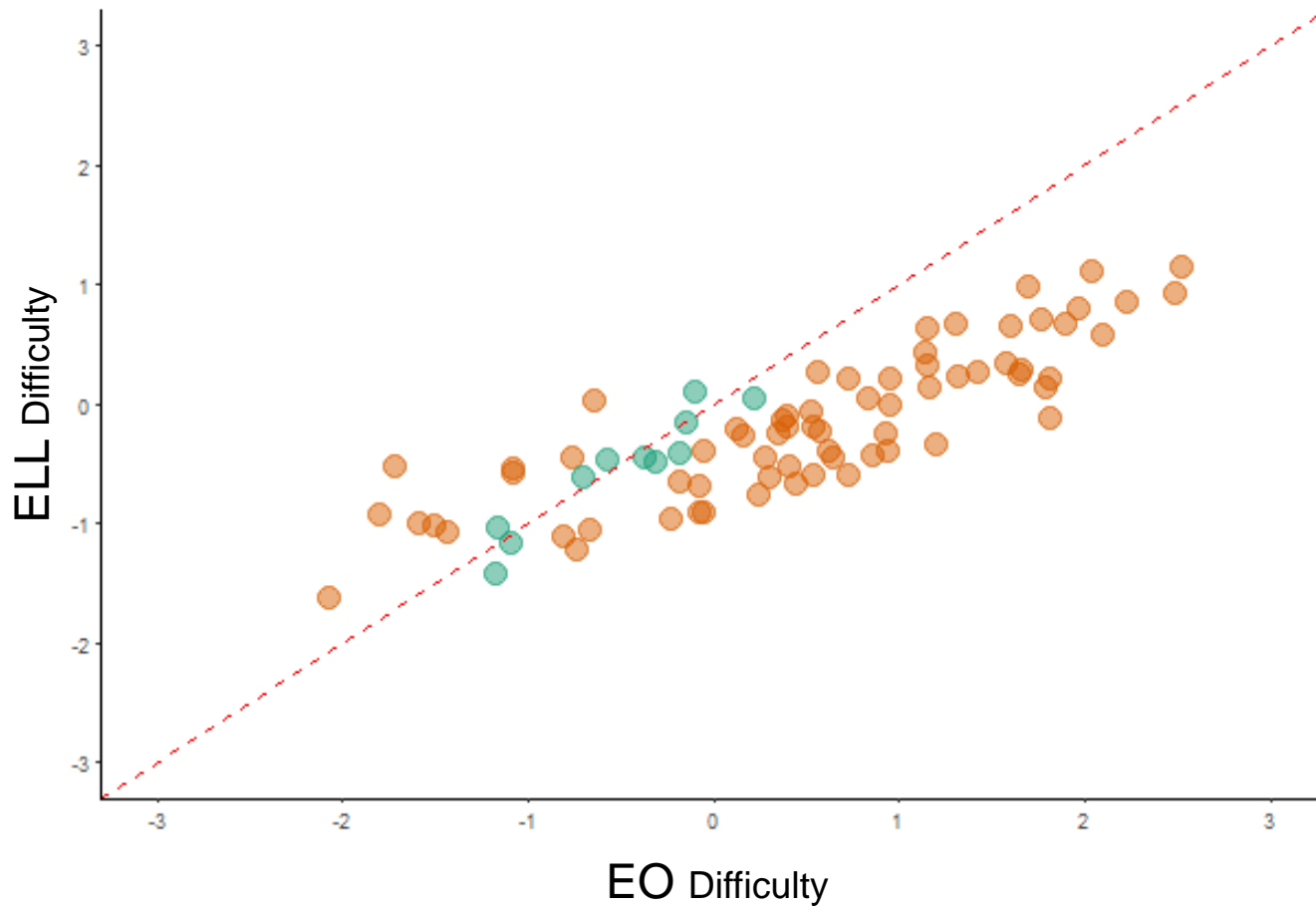


# IRT Approach

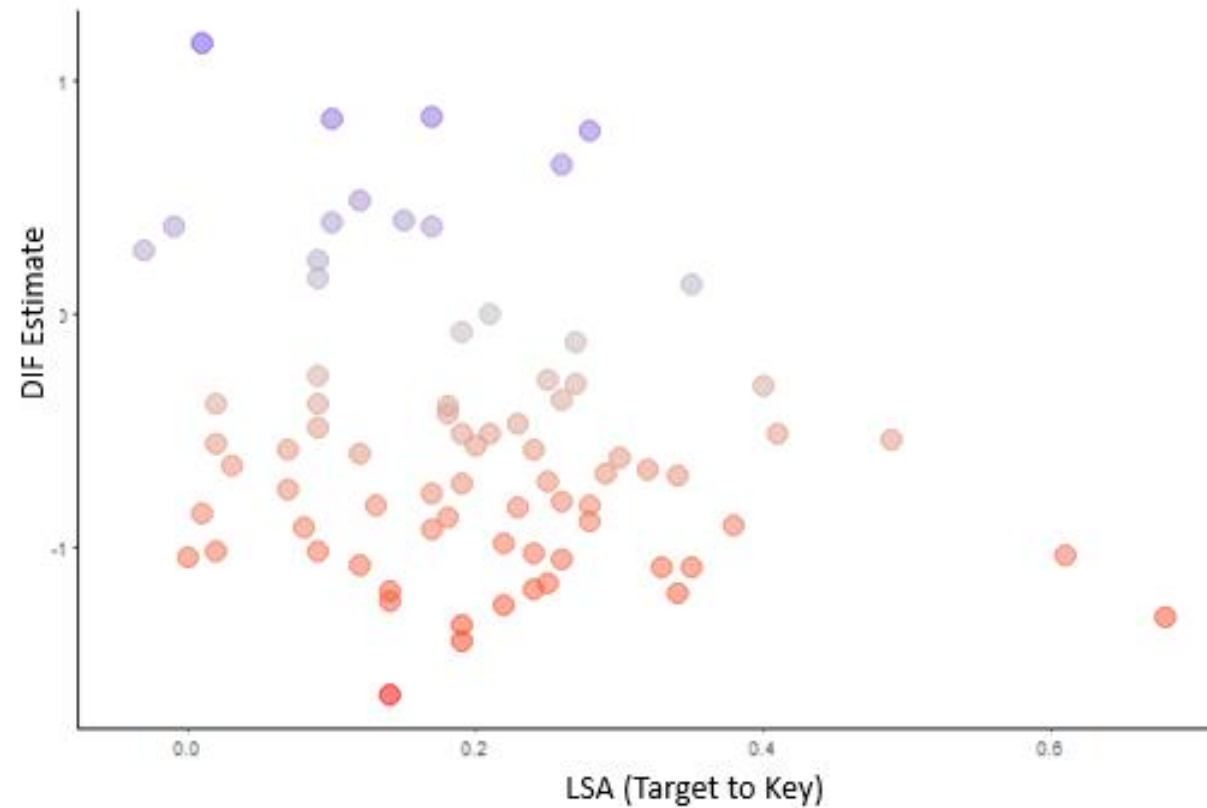
---

- ❑ Logistic regression findings were used to choose anchor items in the IRT Approach
- ❑ 1PL IRT models were used to examine uniform DIF
- ❑ Item characteristics were used to explain difficulty parameters and DIF parameters estimated using 1PL IRT models

# Uniform DIF Results (IRT Approach)



# Relations between DIF and LSA (Preliminary Findings)



Correlation = -0.246

## 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.

---

Thank you for attending!

Paulina.Kulesz@times.uh.edu

The logo for 'TIMES' features the word in a bold, black, serif font. A red, stylized circular graphic element is positioned to the left of the 'T', partially overlapping it.