IDENTIFYING SOURCES OF DIFFERENTIAL ITEM FUNCTIONING ON AN ENGLISH LANGUAGE PROFICIENCY ASSESSMENT

Amy Clark, Neal Kingston

ducational Psychology and Research The University of Kansas

BACKGROUND

METHODS

ELL Students

- During the 2008-2009 school year approximately 5,350,000 ELL students were enrolled in public schools
- More than 400 language groups, with Spanish and Vietnamese as the largest
- 51% increase in ELL students from 10 years prior
- Reauthorization of ESEA in 2001 mandated assessment of ELL students



DIF in Language Accountability Assessments

- •Limited DIF analyses comparing language groups
 - Spanish vs. Non-Spanish
 Gender Groups
 - Gender Groups
 Small sample sizes
 - Small sample sizes

DIF in Language Assessments

Similarities between native and second language (Allalouf & Abramzon, 2007)
 Cognates (Chen & Henning, 1985)
 Access curriculum differently (Abbott, 2007)

Challenge of determining why DIF occurs

•Use of content experts

•Use of effect size to determine extent of DIF (Zumbo, 1999)

Participants

licipants				
		2 nd grade	3 rd grade	Total
	Spanish	16,017	15,359	31,376
	Vietnamese	694	611	1305
	Total	16,711	15,970	32,681

Test Material

- Kansas English Language Proficiency Assessment (Peyton, et al., 2007)
 - o Reading subsection had most items and was self-administered
 - o Administered 2007 2011
- Two forms (Form A 22 items, Form B 23 items)

Procedure

- 1. Items coded for item characteristics
 - Cognates, words with suffixes, multisyllabic words, unique to English sounds from Vietnamese and Spanish (Tang, 2006; Chen & Henning, 1985)
- 2. Logistic regression used to analyze 45 items for uniform and nonunifrom DIF
- 3. Effect size variable created using change in pseudo R²
- 4. Correlation between item characteristics & effect size
- 5. Linear regression with significant characteristics





RESULTS

Uniform DIF

•23 total items with significant group term after controlling for proficiency •12 items favored Vietnamese-speaking students

Nonuniform DIF

•11 total items with significant interaction term after controlling for proficiency and group

• 10 items favored Vietnamese-speaking students

Correlations between Item Characteristics and Effect Size Post Hoc Logistic Regression

•Determine how well item characteristics predicted the occurrence of uniform DIF on an item

- •Classification Accuracy:
- o No predictors: 51%
- o 3 item characteristics: 78%
- Sounds only: 82%

	Effect_size	Multisyllabic	Spanish_sounds	Vietnamese_sounds	Cognate
Multisyllabic	35*				
Spanish_sounds	34*	.72**			
Vietnamese sounds	40**	.74**	.95**		
Cognates	27	.44**	.23	.24	
Suffixes	27	.85**	.66**	.66**	.38**

CONCLUSIONS



Abbott, M. L. (2007). A confirmatory approach to differential item functioning on an ESL reading assessment. Language Testing, 24(1), 7-36. Allalout, A., & Abramzon, A. (2008). Constructing better second language assessments based on differential item functioning analysis. Language Assessment Quarterly, 5(2), 120-141. Chen, Z., & Henning, G. (1985). Linguistic and cultural bias in language proficiency tests. Language Testing, 2(1): 155-163. National Clearinghouse for English Language Acquisition. (2011). The growing number English learner students 198-99 – 2008-09. Washington, DC. Peyton, V., Kingston, N.M., Skorupski, W., Glasnapp, D., & Poggio, J. (2009). Kansas English language proficiency assessment technical manual. Center for Educational Testing and Evaluation, University of Kansas. Tang, G. (2006). Cross-linguistic analysis of Vietnamese and English with implications for Vietnamese language acquisition and maintenance in the United States. Journal of Southeast Asian-American Education & Advancement, 2, 1-33. Zumbo, B. D. (1999). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa, Canada: Directorate of Human Resources Research and Explicition. Denotment of National Defence