Examining Differential Item Functioning from a Latent Class Perspective

Karen Samuelsen
University of Maryland

Differential Item Functioning

DIF occurs when examinees matched on ability have differing probabilities of success on an item.

- Multidimensionality (Ackerman, 1992)
- Compromises our validity argument
- Raises issues of fairness and equity
- Signals the possibility of bias

Problems with Manifest Approaches

Current approaches like Mantel Haenszel, logistic regression, SIBTEST, etc. make comparisons between **manifest** groups

- Gender
- Racial groups
- Ethnic groups

Problems with Manifest Approaches

Problems:

- Manifest groups are not homogeneous (Cohen & Bolt, 2002)
- Interactions are really where the action is (Hu & Dorans, 1989)
- Manifest groups are proxies for an "educational advantage attribute"

What if we use manifest groups instead of latent ones?

1. We incorrectly assume items exhibiting DIF impact all members of a manifest group
2. Miss items functioning differentially based on the latent attribute but not the manifest
3. Underestimate the magnitude of the 'true DIF'

Steps in this Research

1. Simulation study using the MH procedure to make the case that manifest approaches to DIF are problematic
2. Using the simulated data, examine the efficacy of using the Mixed Rasch Model with WINBUGS

Steps in this Research

3. Develop a series of protocols for examining differential item functioning using a latent class approach
4. Use these protocols in the analysis of a test of English language proficiency

Simulation Study

Data simulated for a 20-item test and 6 factors were manipulated:

- Sample size (500 and 2000 examinees)
- Overlap between manifest groups and latent proportions (60%, 70%, 80%, 90%, 100%)
- Manifest proportions (50/50, 80/20)

Simulation Study

Other factors manipulated:

- Number of items exhibiting DIF (2, 6, or 10)
- Effect size (Δb = 0.4, 0.8 or 1.2)
- Ability distributions (Normal(0,1) or (-1,1))

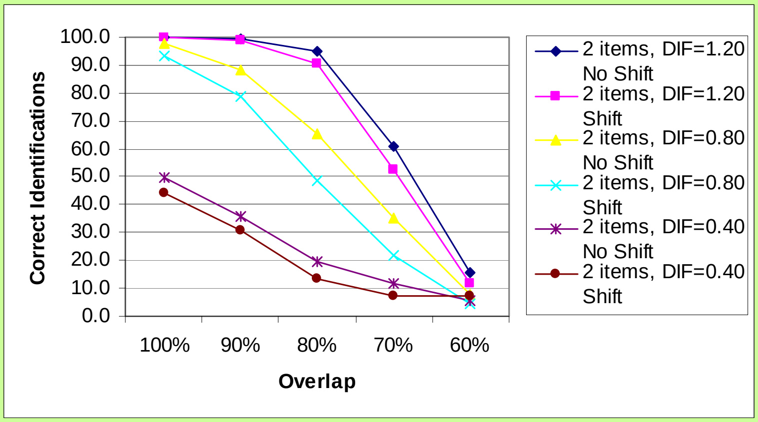
Data simulated using a GAUSS program and analyzed using MH procedure

Simulation Study

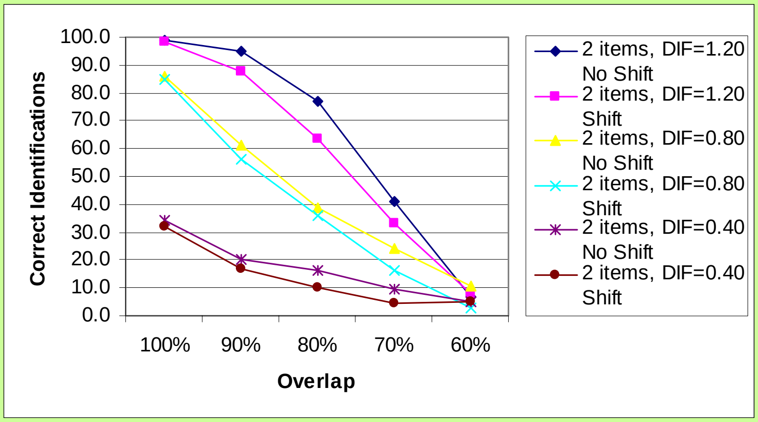
How are each of following impacted by manipulating the factors:

- Power to detect differential functioning
- Magnitude of the DIF
- Type 1 error rate

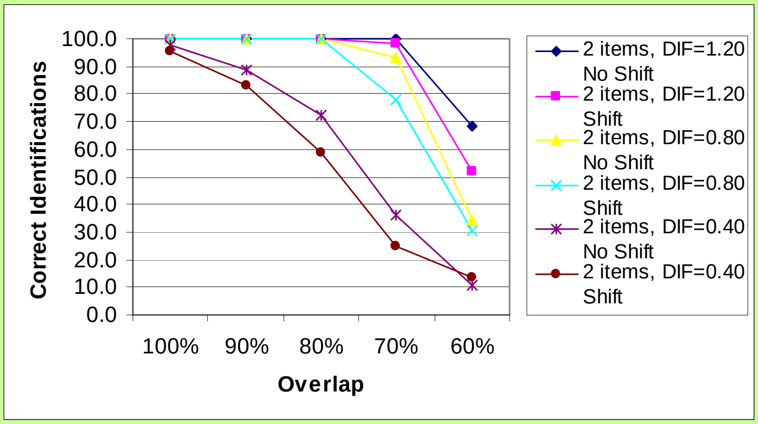
Correct identifications
50/50 split with 500 examinees



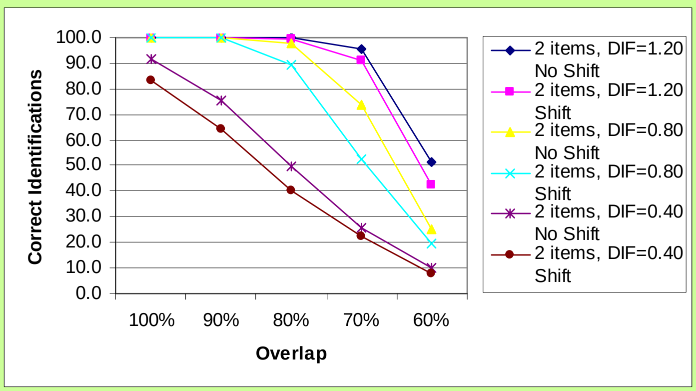
Correct identifications
80/20 split with 500 examinees



Correct identifications 50/50 split with 2000 examinees



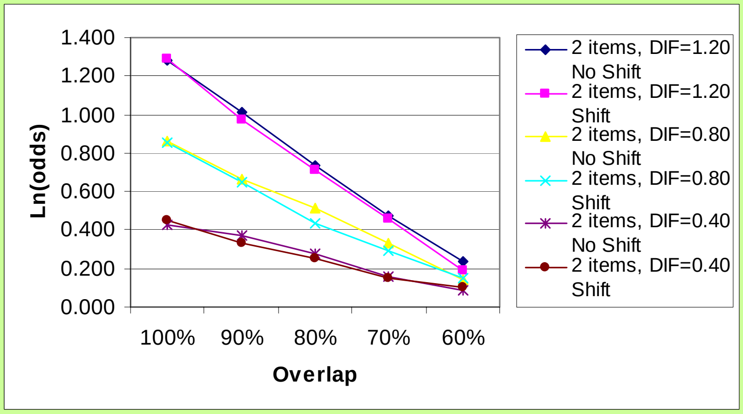
Correct identifications 80/20 split with 2000 examinees



Overlap Necessary for Power > 0.80

	2000		500	
	DIF	Overlap	DIF	Overlap
50/50	1.20	0.70	1.20	0.80
	0.80	0.80	0.80	0.90
	0.40	1.00	0.40	Never
80/20	1.20	0.70	1.20	0.90
	0.80	0.80	0.80	1.00
	0.40	1.00	0.40	Never

Ln(odds) as a Function of Overlap



Overlap necessary for to Escape B or C Classification*

B Classification		C Classification	
Magnitude	Overlap	Magnitude	Overlap
1.20	70%	1.20	80%
0.80	80%	0.80	90%
0.40	100%	0.40	Never

* A, B and C Classifications used by ETS (Zieky, 1993)

Misclassifications or Type 1 errors

Regression analyses showed that the following were significant predictors of Type 1 errors

- Sample size
- Contamination of the matching criterion
- Degree of overlap
- Manifest proportions

Mixed Rasch Model

The Rasch model can be used to describe "the response behavior of all persons within a latent class, but that different sets of item parameters hold for the different latent classes" (Rost, 1990)

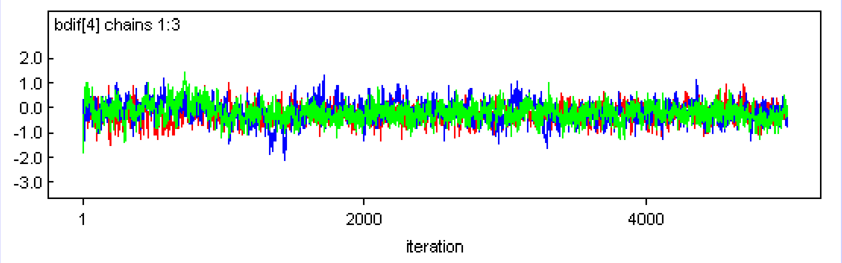
$$p_{ni} = \sum_g \pi_g \frac{\exp(\theta_{ng} - b_{ig})}{1 + \exp(\theta_{ng} - b_{ig})}$$

Findings from Recovery Study

1. All items with DIF correctly identified for all overlap conditions (for n=2000)
2. Magnitude of DIF well estimated for easy items
3. Manifest proportions -- OK
4. Means of ability distributions -- OK

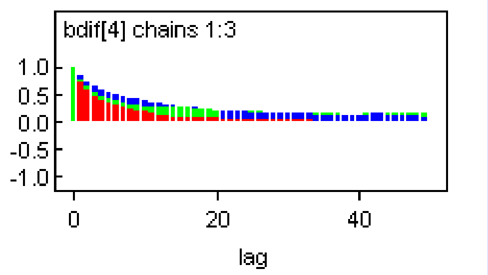
History Plot for bdif

Difference in item difficulties for 500 examinees



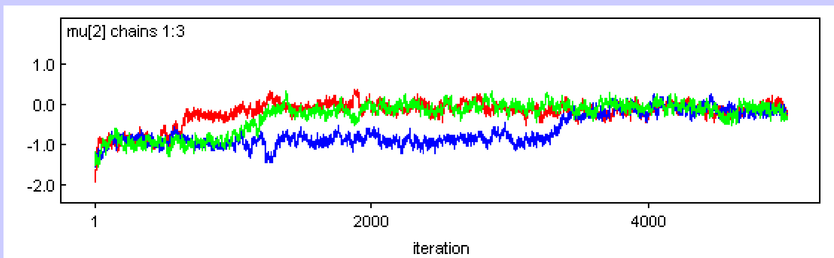
Autocorrelations for bdif

Difference in item difficulties for 500 examinees



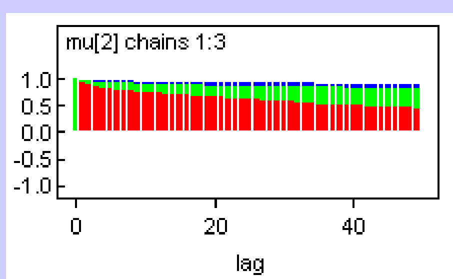
History Plots for mu

Means of latent ability distributions for 500 examinees



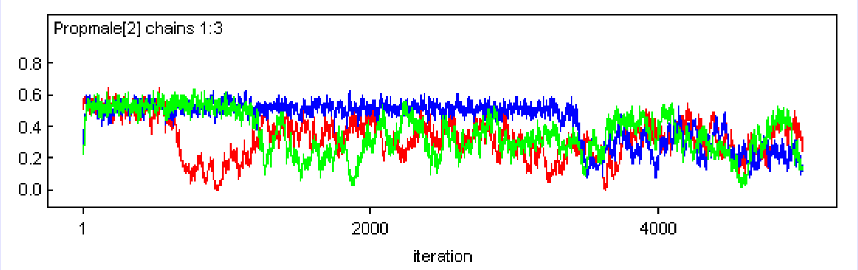
Autocorrelations for mu

Means of latent ability distributions for 500 examinees



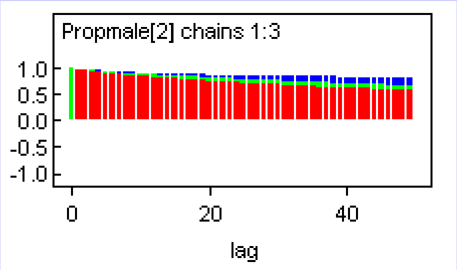
History Plots for Proportions

Proportion of males in the 2nd latent class for 500 examinees



Autocorrelations for Proportions

Proportion of males in the 2nd latent class for 500 examinees



Four Step Approach

1. Identify the model that best fits the data;
2. Decide whether the manifest group percentages within the latent classes warrant a latent class approach;
3. Examine the data from the latent class analysis for clues as to why there is DIF and to inform the choice of covariates;
4. Use the covariates to predict membership in the latent classes.

Sample from ELDA

1016 Students

- Males and females
- Asian and Hispanic students representing a variety of countries
- 3rd, 4th and 5th graders
- Some ELL's born in the US

More info about the sample

34 multiple-choice items were used:

Mantel Haenszel procedure results

- Items 18, 25, 30 and 34 exhibited DIF with respect to ethnicity
- Items 7, 9, 23, 27, 33 and 34 showed gender DIF

Checking Model Fit

Shadow data technique results for fit

	Mean	SD	2.5%	Median	92.5%
1 class	0.489	0.015	0.459	0.489	0.520
2 class	0.516	0.015	0.4865	0.516	0.546
3 class	0.555	0.014	0.527	0.555	0.583

Results of latent class DIF analysis

First latent class

- 90.8% of the Asian females
- 74.9% of the Asian males

83%

- 82.0% of Hispanic females
- 64.9% of Hispanic males

74%

Thus, 83% of females and 66% of males

Results of latent class DIF analysis

- Examinees in the first class were on average much more able than those in the second class.
- 23 of the 34 test items exhibited statistically significant DIF
- The items found using MH were a subset of the items found using this latent analysis

Determining why the DIF occurs

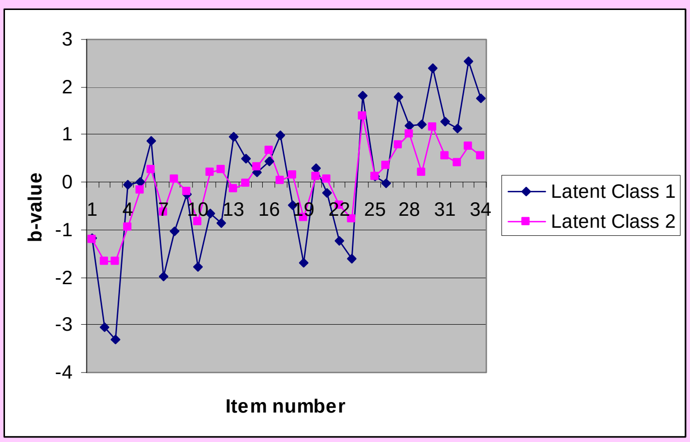
Look at the features of the items

- Short passages, Instructions, Long passages
- Level of cognitive thinking

Noteworthy evidence

- Items 29-34 which all refer to one reading passage showed DIF

Results of latent class DIF analysis



Results of latent class DIF analysis

Use covariates to predict latent class membership

- Birth country (US or not)
- Type of instruction
- Years of ESL instruction

Implications of this Research

1. Real data showed an overlap condition is more problematic that shown in the simulation study
2. Sample sizes currently considered acceptable are too low.
3. DIF uncovered by traditional approaches may be attributable to differences in small numbers of examinees

Next Steps

Use more complex models

- 2-PLM, 3-PLM
- Incorporate guessing
- Elemental components of item difficulty

Apply to other types of tests (especially achievement tests)

Concerns and Conclusions

A latent class approach is more difficult than the manifest approaches currently used.

- The manifest approaches are politically expedient
- They yield results that are easy to understand.

Concerns and Conclusions

A latent class approach can yield information that is more accurate and enlightening

- Items with DIF
- Proportions of manifest groups in classes
- Means of latent classes