

Investigating Item Bias in a CS1 Exam with Differential Item Functioning

SIGCSE 2021

Matt J. Davidson

Brett Wortzman, Amy J. Ko, Min Li

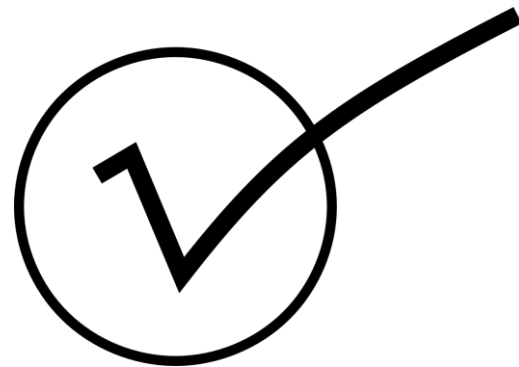


Impacts of testing



What makes a good assessment?

- Validity
 - measures what we intend
- Reliability
 - measures consistently
- Fairness
 - measures *only* what we intend



Why is fairness so important?

- Academic integrity
- Validity



**How can we be sure our exams
are unbiased?**

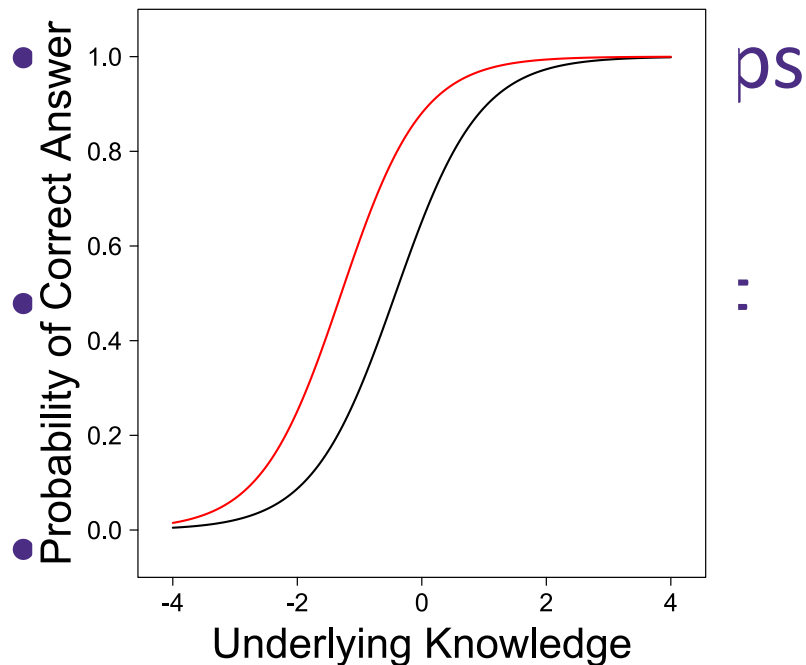
DIF is a method to investigate fairness.

~~Does the test work similarly for all students?~~

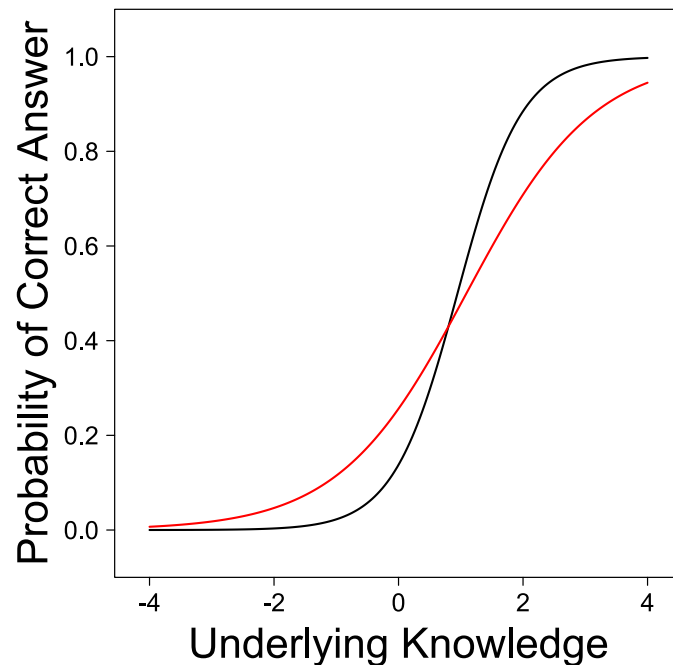
Did students with similar knowledge perform similarly?

What can DIF show us?

Uniform DIF

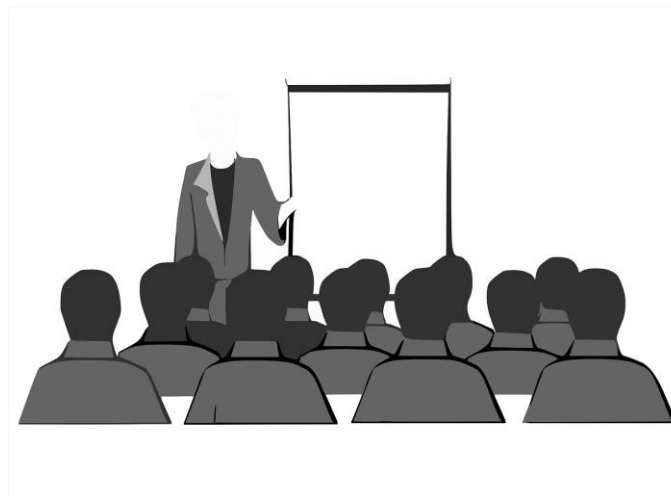


Non-uniform DIF



How can I use DIF methods?

- Instructors
 - Grading
 - Item banks
- Researchers
 - Validation
 - Study designs

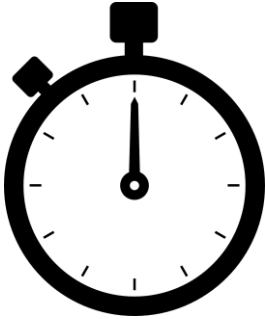


Basic DIF Procedure

- Check response data
- Apply DIF methods
- Examine effect sizes
- Remove or revise items



CS1 Final Exam



Basic Psychometric Properties

	Difficulty
RefMyst	0.65
ArraySim	0.66
InheritMys	0.71
switch1	0.28
switch2	0.45
filter	0.82
isFiblike	0.53
Critters	0.45
delta	0.35
numWord	0.33

Our analysis

- Sample size: 939 students
- Binary gender
 - 360 reported female
- Year of study
 - 246 beyond first year







DIF methods make comparisons.

- Compare model fits
 - Logistic regression
 - Item Response Theory
- Matching
 - Total score
 - Underlying knowledge



Did we find any DIF items?

	Binary Gender	Year of study
Logistic Regression		
Likelihood Ratio Test		

Investigating Item Bias in a CS1 Exam using Differential Item Functioning

Matt J. Davidson

mattjd@uw.edu
mattjohndavidson.github.io
@dontmattme

Brett Wortzman

brettwo@cs.uw.edu

Amy J. Ko

ajko@uw.edu

Min Li

minli@uw.edu

Main takeaway: We must start using DIF methods if we care about fairness and validity in our exams.

Want to learn more? Blog post with details, example analysis, and R code: **bitly.com/sigcse21DIF**