Is the Force Concept Inventory Biased? Investigating Differential Item Functioning on a Test of Conceptual Learning in Physics

Sharon E. Osborn Popp, David E. Meltzer, and
Colleen Megowan-Romanowicz
Arizona State University

Paper Presented at the Annual Meeting of the
American Educational Research Association
New Orleans, LA April, 2011

ASU Mary Lou Fulton Teachers College
ARIZONA STATE UNIVERSITY

Overview

- Examined possible gender bias on a widely-used measure of conceptual knowledge in physics

- A Differential Item Functioning (DIF) analysis was conducted on 4775 responses to the Force Concept Inventory (FCI)

Background: The Force Concept Inventory

- **First published in *The Physics Teacher*, 1992**
  - Hestenes, Wells, & Swackhamer
- **Revised 1995 – minor changes; scores comparable***
  - Halloun, Hake, Mosca, & Hestenes
- **30 MC items – intended to assess basic concepts of force and kinematics**
- **Most widely-used measure of mechanics concepts by physics educators and researchers**
- **Translated into 18 languages**

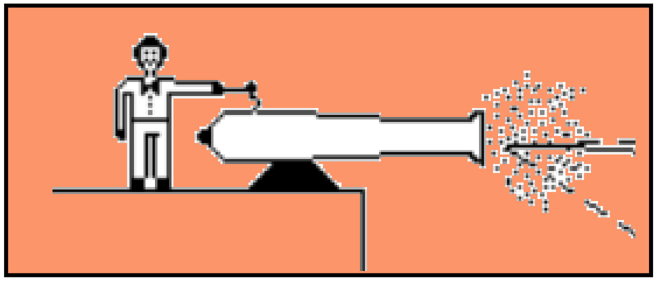
***Note: 27 of 30 items on FCI-REV95 are the same or similar to FCI92, but the items are ordered differently

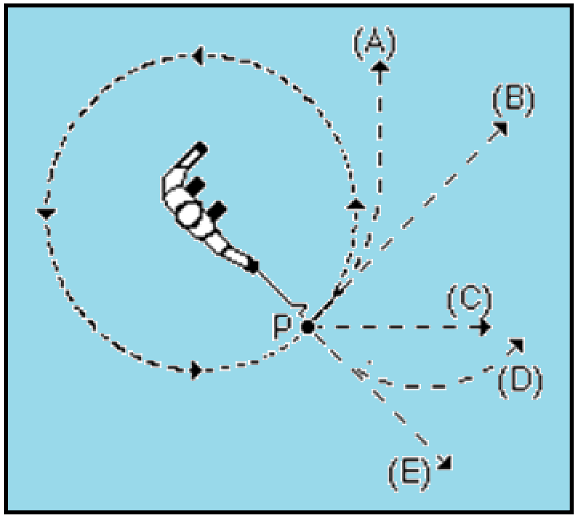
Background: The Physics Gender Gap

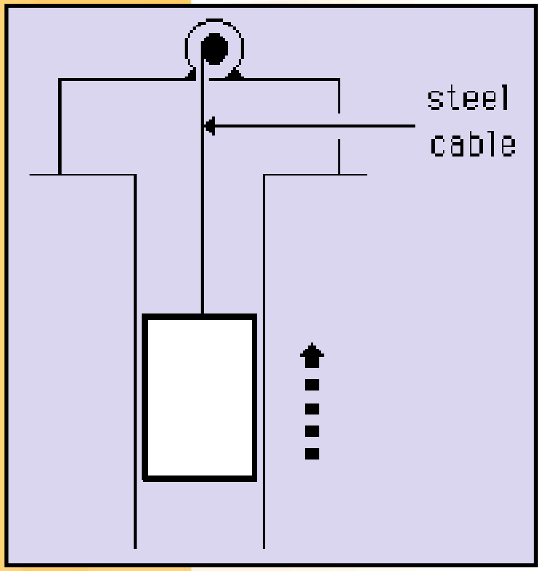
- **Persistent differences between females and males in performance on measures of conceptual knowledge in science/physics**
  - E.g., Kahle & Meece (1994) and
  - Mullis, Martin, Fierros, Goldberg, & Stemler (2000)
- **Attempts to explain or reduce the gap via background variables/instructional intervention have been mixed**
  - E.g., Lorenzo, Crouch, and Mazur (2006),
  - Pollock, Finkelstein, and Kost (2007),
  - Kost, Pollock, & Finkelstein (2009), and
  - Miyake, Kost-Smith, Finkelstein, Pollock, Cohen, and Ito (2010)

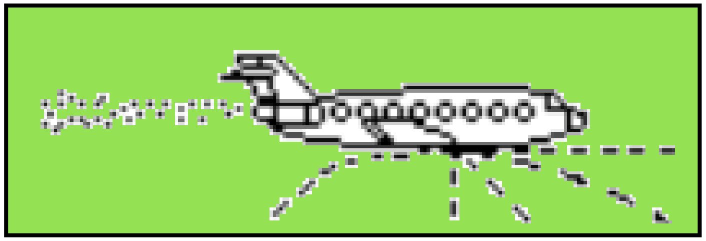
Could differences between males and females be due to test bias?

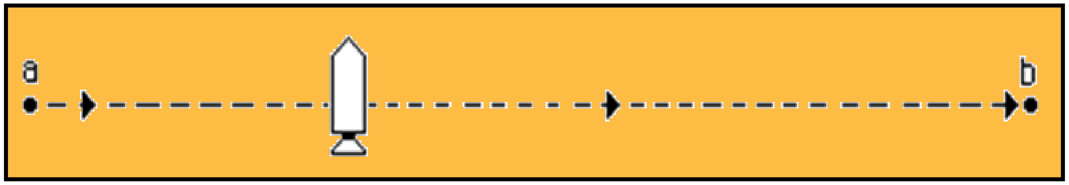
- Concerns raised that properties of the FCI itself, unrelated to student ability, influence performance
- Situational contexts seem male-oriented and lab-oriented (e.g., rockets, cannons, steel balls)











Possible FCI Bias?

- **McCullough & Meltzer, 2001**
  - Females had much higher rate of correct response on items 14 and 23 on a female-context version of FCI
- **McCullough, 2004**
  - Males performed less well on the female-context version; however, females did not perform significantly better, overall
- **Docktor & Heller, 2008**
  - Items 14 and 23 had largest male-female differences in correct response on standard FCI

Purpose: Investigate Possible Bias on the FCI

- Systematic item bias can weaken inferences or even mislead

- Educators and researchers need to have confidence in measurement instruments

- Detection of Differential Item Functioning can reveal possible bias

Differential Item Functioning (DIF)

“Differential item functioning exists when examinees of equal ability differ, on average, according to their group membership in their particular responses to an item” (p. 81)

The Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999)

Differential Item Functioning (DIF)

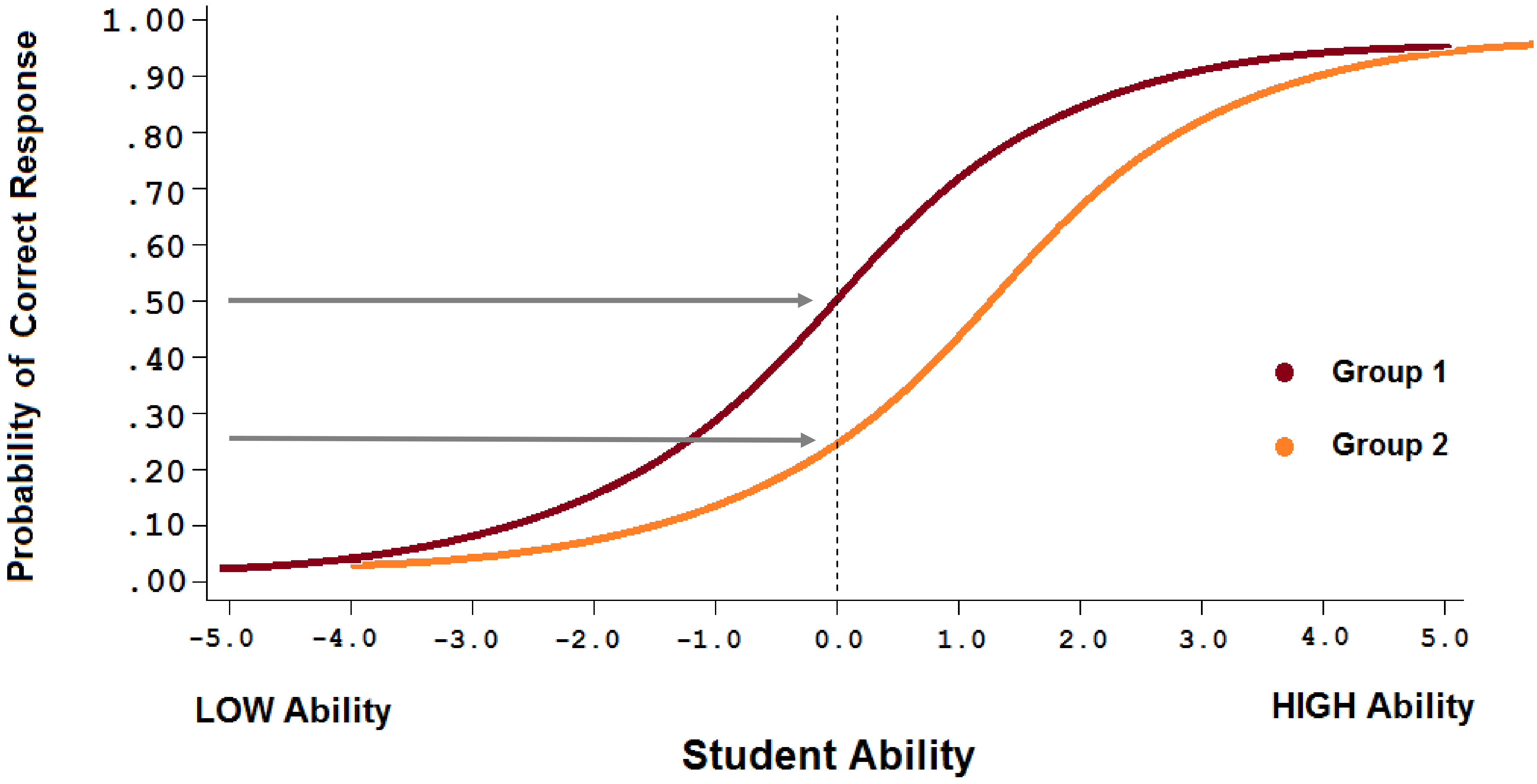
- DIF is present when students at the ***same ability*** level show unexpectedly different performance on a given test item

- DIF methods have evolved over the years and have become a standard part of large-scale assessment programs

- The presence of DIF does not necessarily mean an item is biased; judgmental review is essential to confirm bias

Item Characteristic Curves: Hypothetical Item exhibiting DIF

Same ability, but probability of responding correctly is .50 for Group 1 and .25 for Group 2



Sample

- **4775 high school physics students**
  - Regular and honors physics
  - Collected over four years from teachers around the US*
  - Took FCI as a posttest following completion of mechanics curriculum
  - **2348 Females (49%)**
  - **2427 Males (51%)**

* Data collected during the course of Modeling Instruction in Physics Workshops, sponsored by the NSF

Method

Rasch (one-parameter logistic IRT) model

- Probabilistic model
- The probability (P) of a correct response, given ability (b) and difficulty (d) is given by:

$$\mathbf{P\ (b,\ d) =} \quad \frac{e^{(b-d)}}{1 + e^{(b-d)}}$$

Where: **e = 2.718 (base of the natural log system)**

b = student’s ability

d = item’s difficulty

DIF Analysis

- **Rasch Model Requirements (c.f., assumptions)**
  - Estimates of item difficulty must be invariant across different samples from the same population
- **DIF Contrast Value**
  - The difference between an item's difficulty estimates for females and males
- ***t*-tests to assess differences**
  - Are routinely computed, but are not considered appropriate as a measure of practical DIF (Camilli, 2006)

- **The DIF Contrast logit value provides an appropriate effect size (Wang, 2009)**

- **DIF values of .50 logits used as cut-off for substantial DIF**

Results: Raw Scores

Females	All	Males
13.52	15.63	17.66
(*SD* = 5.96)	(*SD* = 6.74)	(*SD* = 6.82)

Correlation between proportions correct between Females and Males was .89

Males had a higher proportion correct for all items, with differences ranging from .03 to .28

Results: Rasch Parameter Estimates (in logits)

Females	All	Males
-0.24	**0.19**	**0.61**
(*SD* = 1.13)	**(*SD* = 1.37)**	**(*SD* = 1.49)**

Correlation between item difficulty estimates between Females and Males was .89

- DIF contrast values were 0 for 5 items
- DIF contrast values for 14 items had significant *t*

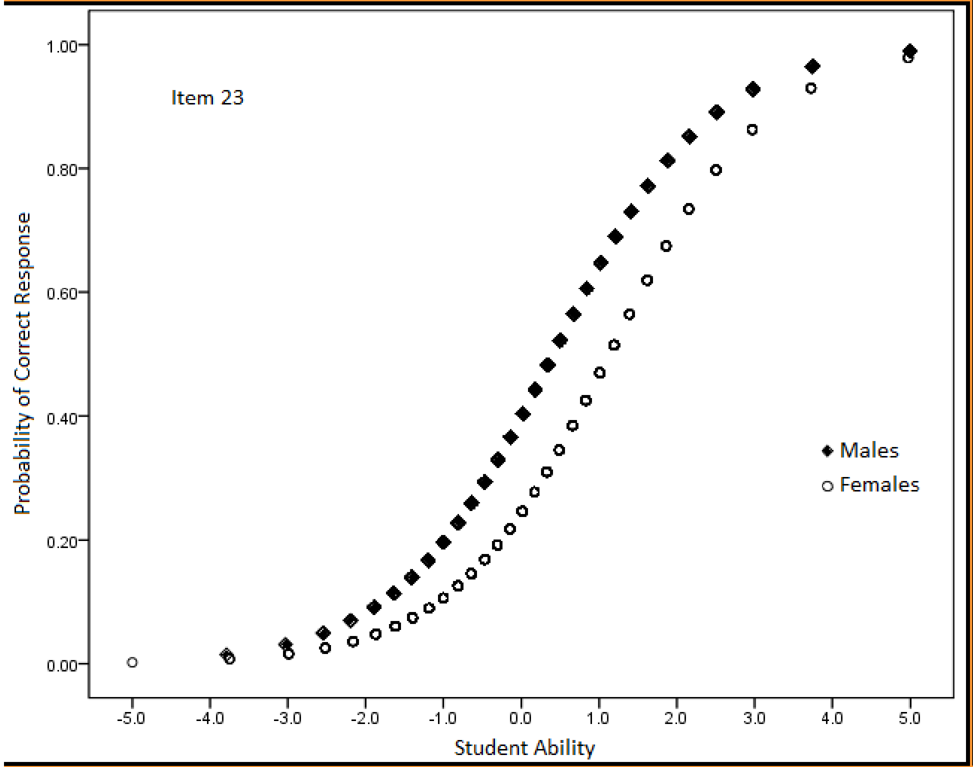
  7 favored Males; 7 favored Females
- DIF contrast values for 3 items exhibited substantial DIF (i.e., contrast exceeded .50 logits)

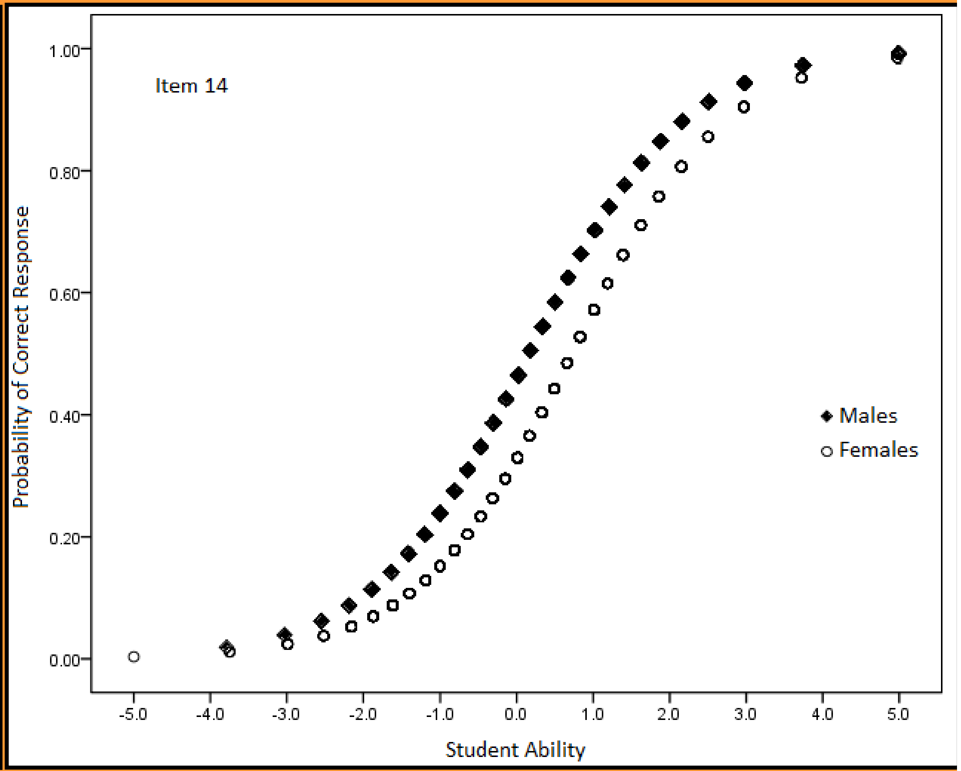
  2 favored Males; 1 favored Females

FCI Items Exhibiting Substantial DIF

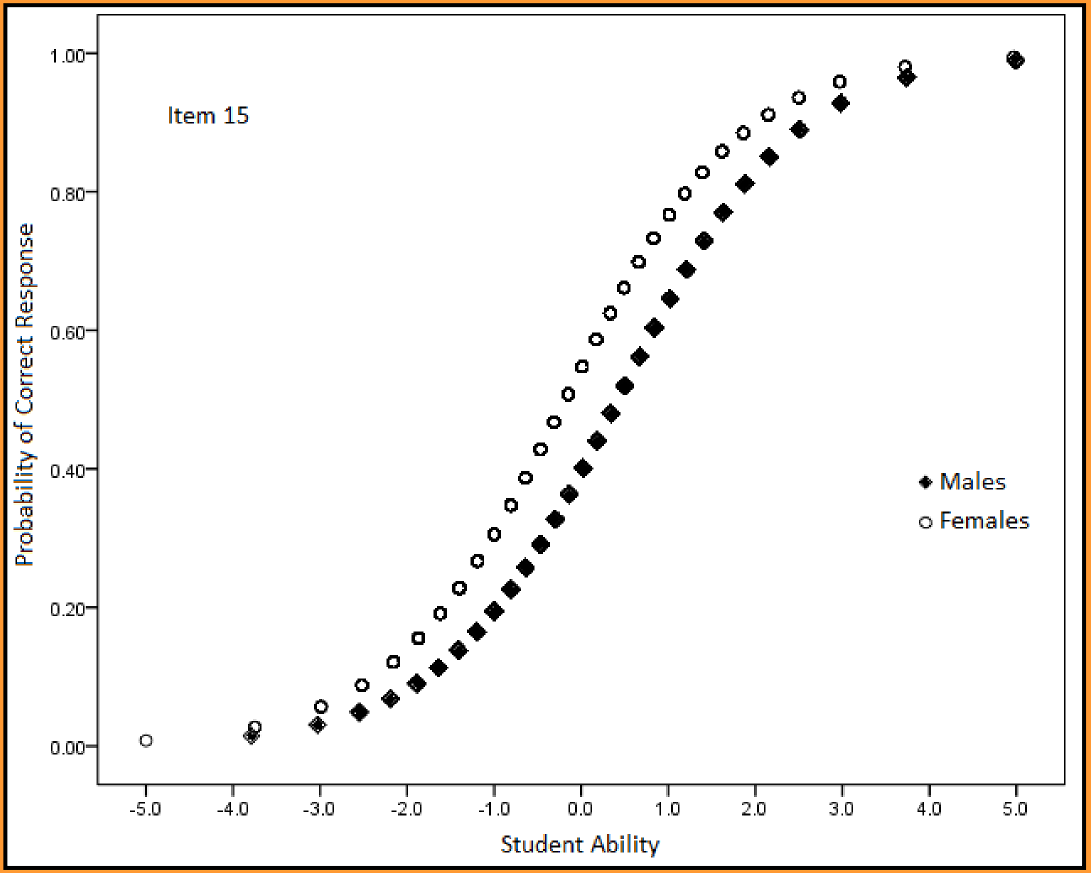
- 2 Favored Males (positive contrast values)
  - Item 23: DIF Contrast Value of 0.73 logits
  - Item 14: DIF Contrast Value of 0.57 logits
- 1 Favored Females (negative contrast values)
  - Item 15: DIF Contrast Value of -0.59 logits

Items 23 and 14 (Favored Males)



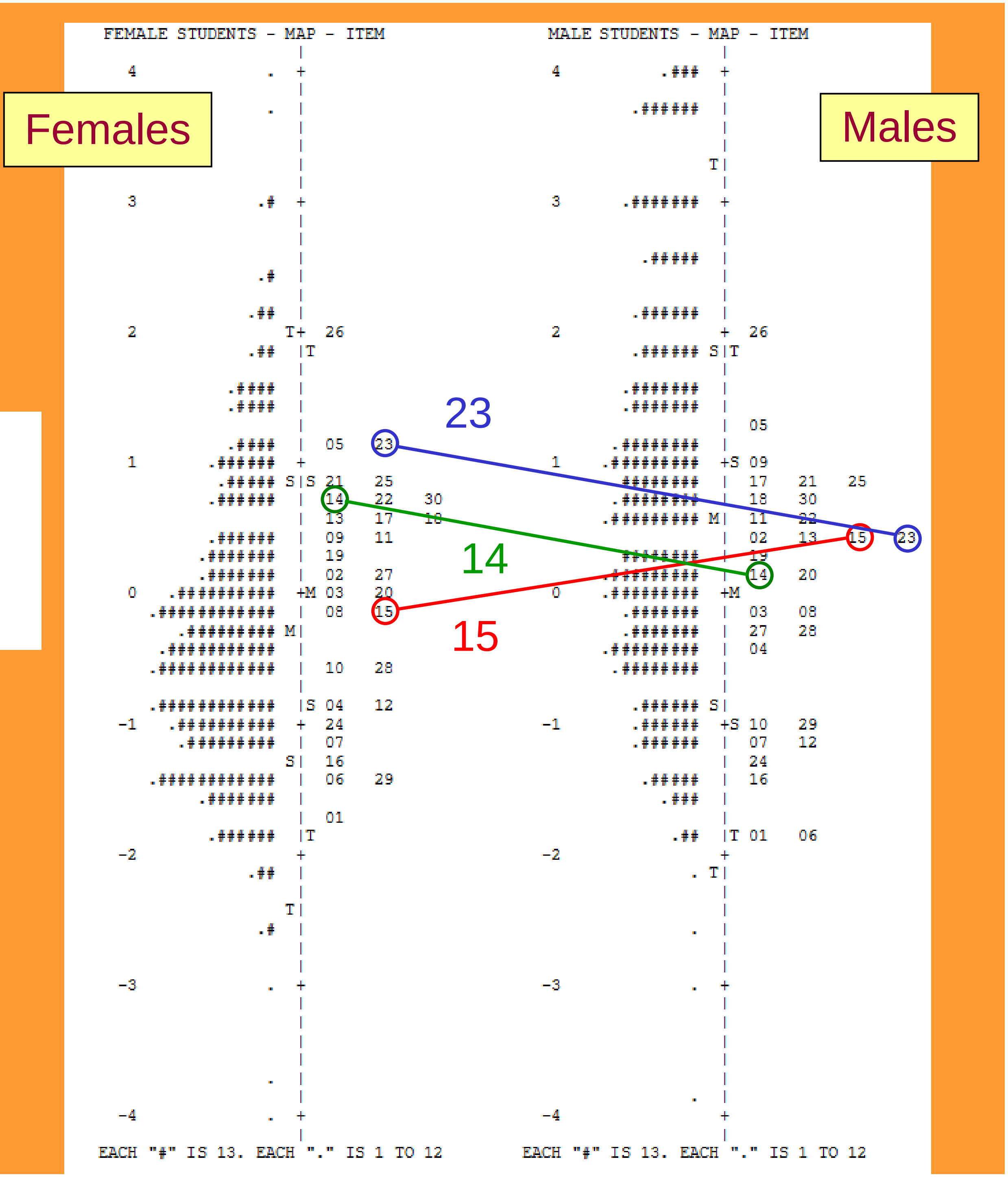


Item 15 (Favored Females)



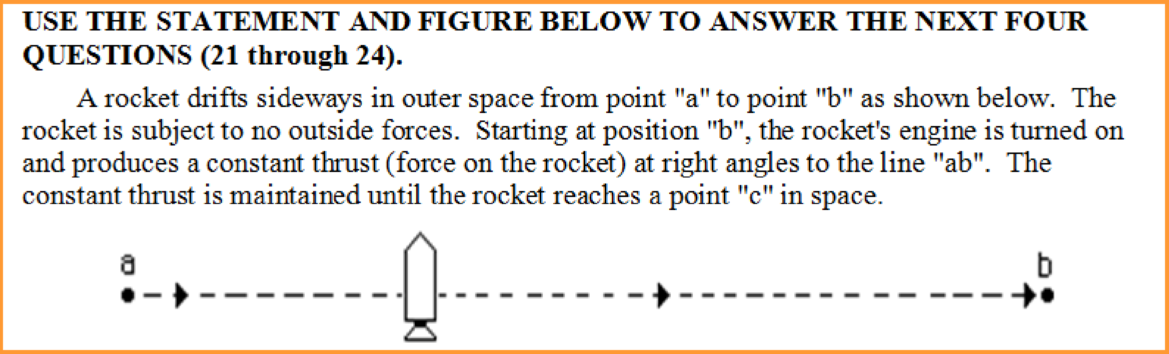
Wright Maps

Females and Males



```
  FEMALE STUDENTS - MAP - ITEM
                     |
    4             .  +
                     |
                  .  |
                     |
                     |
                     |
                     |
    3            .#  +
                     |
                     |
                     |
                 .#  |
                     |
                .##  |
    2               T+  26
                .##  |T
                     |
              .####  |
              .####  |
                     |
              .####  |  05    23
    1       .######  +
             .##### S|S 21    25
            .######  |  14    22    30
                     |  13    17    18
            .######  |  09    11
           .#######  |  19
           .#######  |  02    27
    0   .##########  +M 03    20
      .############  |  08    15
         .######### M|
       .###########  |
      .############  |  10    28
                     |
      .############  |S 04    12
   -1   .##########  +  24
         .#########  |  07
                    S|  16
      .############  |  06    29
           .#######  |
                     |  01
            .######  |T
   -2                +
                .##  |
                     |
                    T|
                 .#  |
                     |
                     |
   -3             .  +
                     |
                     |
                     |
                     |
                  .  |
                     |
   -4             .  +
                     |
EACH "#" IS 13. EACH "." IS 1 TO 12
```

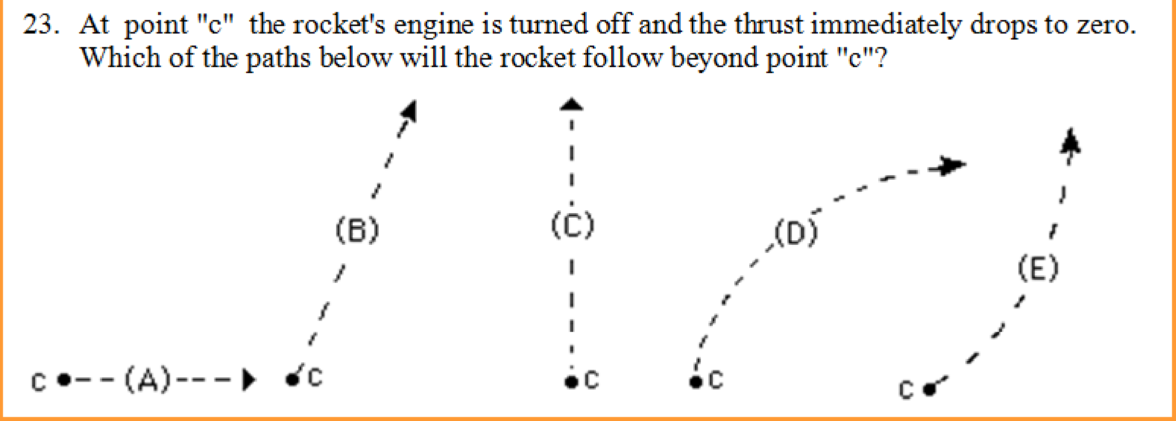
```
  MALE STUDENTS - MAP - ITEM
                     |
    4          .###  +
                     |
            .######  |
                     |
                     |
                    T|
                     |
    3      .#######  +
                     |
                     |
             .#####  |
                     |
                     |
            .######  |
    2                +  26
            .###### S|T
                     |
           .#######  |
           .#######  |
                     |  05
          .########  |
    1     .#########  +S 09
           ########  |  17    21    25
          .########  |  18    30
         .######### M|  11    22
                     |  02    13    15    23
           ########  |  19
         .#########  |  14    20
    0    .#########  +M
           .#######  |  03    08
           .#######  |  27    28
         .#########  |  04
          .########  |
                     |
            .###### S|
   -1       .######  +S 10    29
            .######  |  07    12
                     |  24
             .#####  |  16
               .###  |
                     |
                .##  |T 01    06
   -2                +
                  . T|
                     |
                     |
                  .  |
                     |
                     |
   -3             .  +
                     |
                     |
                     |
                     |
                     |
                  .  |
   -4                +
                     |
EACH "#" IS 13. EACH "." IS 1 TO 12
```



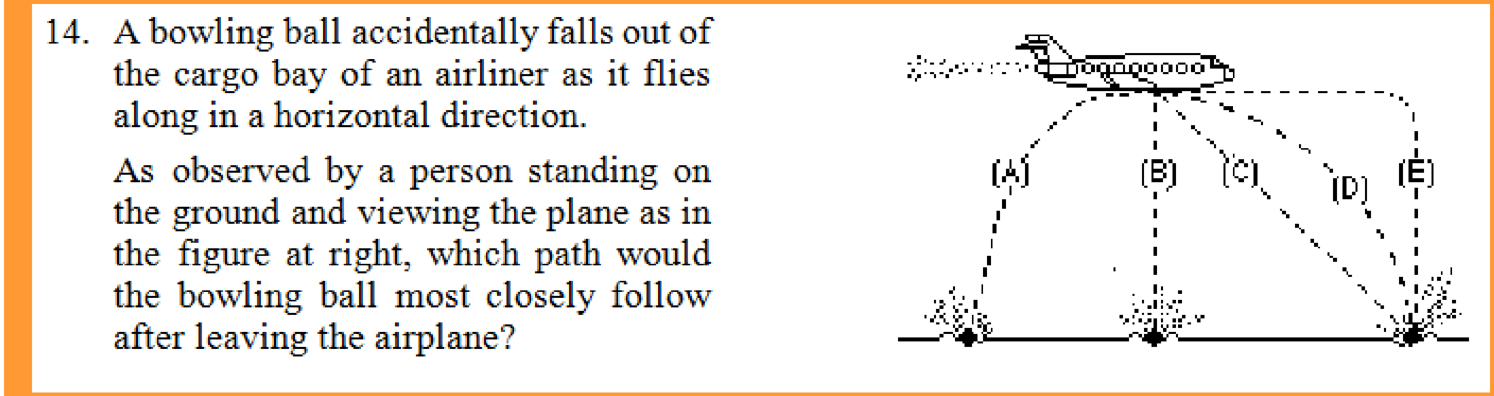
USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT FOUR QUESTIONS (21 through 24).

A rocket drifts sideways in outer space from point "a" to point "b" as shown below. The rocket is subject to no outside forces. Starting at position "b", the rocket's engine is turned on and produces a constant thrust (force on the rocket) at right angles to the line "ab". The constant thrust is maintained until the rocket reaches a point "c" in space.

Item 23 >



23. At point "c" the rocket's engine is turned off and the thrust immediately drops to zero. Which of the paths below will the rocket follow beyond point "c"?



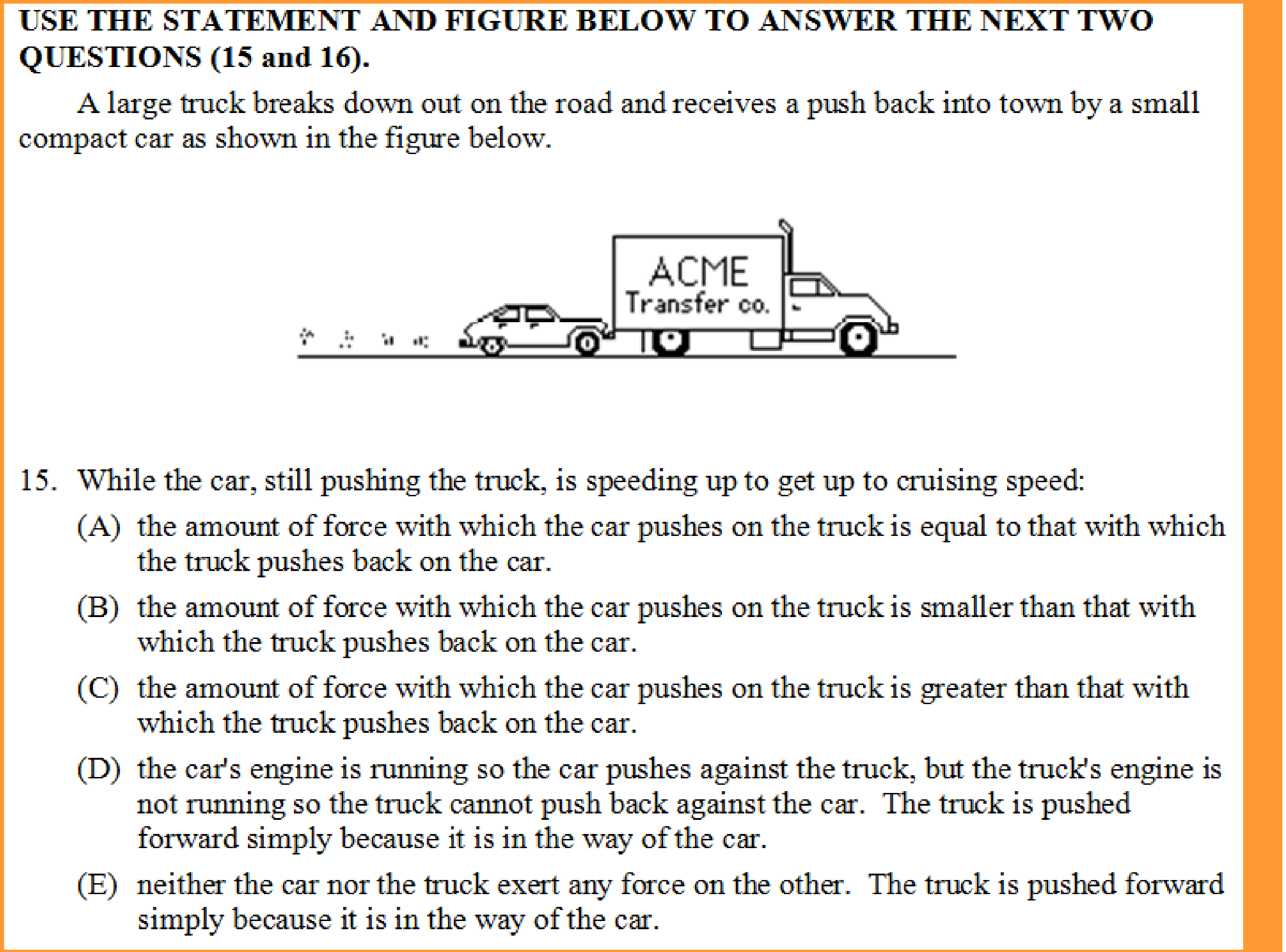
14. A bowling ball accidentally falls out of the cargo bay of an airliner as it flies along in a horizontal direction.

As observed by a person standing on the ground and viewing the plane as in the figure at right, which path would the bowling ball most closely follow after leaving the airplane?

< Item 14

Path prediction, pictorial response; however, others items with similar features did not exhibit DIF

Item 15 >


ACME
Transfer co.

USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT TWO QUESTIONS (15 and 16).

A large truck breaks down out on the road and receives a push back into town by a small compact car as shown in the figure below.

15. While the car, still pushing the truck, is speeding up to get up to cruising speed:

(A) the amount of force with which the car pushes on the truck is equal to that with which the truck pushes back on the car.

(B) the amount of force with which the car pushes on the truck is smaller than that with which the truck pushes back on the car.

(C) the amount of force with which the car pushes on the truck is greater than that with which the truck pushes back on the car.

(D) the car's engine is running so the car pushes against the truck, but the truck's engine is not running so the truck cannot push back against the car. The truck is pushed forward simply because it is in the way of the car.

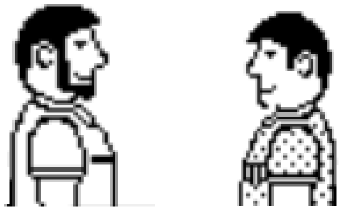
(E) neither the car nor the truck exert any force on the other. The truck is pushed forward simply because it is in the way of the car.

Wordy; however, other items dependent on reading many words did not exhibit DIF

Discussion

- No clear trend favoring males
- Three items exhibited substantial DIF
  - 2 favored Males, 1 favored Females
- Items 23 and 14, favoring males, have been cited previously
  - Docktor & Heller (2006)
  - McCullough & Meltzer (2001)
- No obvious reason for items 23, 15, & 14 to be biased upon review
- Re-scoring without substantial DIF items does not change Male-Female difference in performance (effect size remains moderately large)

Conclusion and Next Steps



- DIF analysis provides some evidence to support the valid use of the FCI for assessment and research
  - Findings suggest that the FCI is not systematically biased in favor of males
- The three items that exhibited substantial DIF warrant continued attention
- Additional DIF analyses on this and other samples are needed to confirm current findings
- Explore the effect of item dependencies (two items, 15 & 23, shared a common context with other items)

For questions or a copy of the paper/powerpoint please contact:

Sharon E. Osborn Popp

osbornpopp@asu.edu