

Introduction

- The Force Concept Inventory (FCI) is an exam normally given in intro physics.
- Traxler *et al.* (2018) examined the structure of the FCI and found that men were significantly advantaged.
- We aim to replicate this study across many institutions and a wider dataset.
- Data were obtained from LASSO: Learning About STEM Student Outcomes, an online service for administering concept inventories to students.

Research Question

- Traxler's question:* Are there FCI items that would be identified as problematic? If so, are the problematic items consistent across gender?
- Our question:* Do we identify similar structural features of the FCI when using a dataset containing multiple institutions?

Methodology

Our Study	Traxler
Online	Paper
Sample size around 5000	Sample size around 4700
Many small courses	Few large courses
More demographic data	Less demographic data

- We filtered for:
 - Time on test < 5 minutes
 - < 80% completed
 - Courses with < 9 students
 - Courses with < 40% participation
- We ran Classical Test Theory (CTT) and Item Response Theory (IRT) to identify problematic items.
- Problematic items are: too difficult, too easy, or do not discriminate (differentiate between high and low performers).
- All analysis was done in R.

Results

Problematic Items Found on the FCI

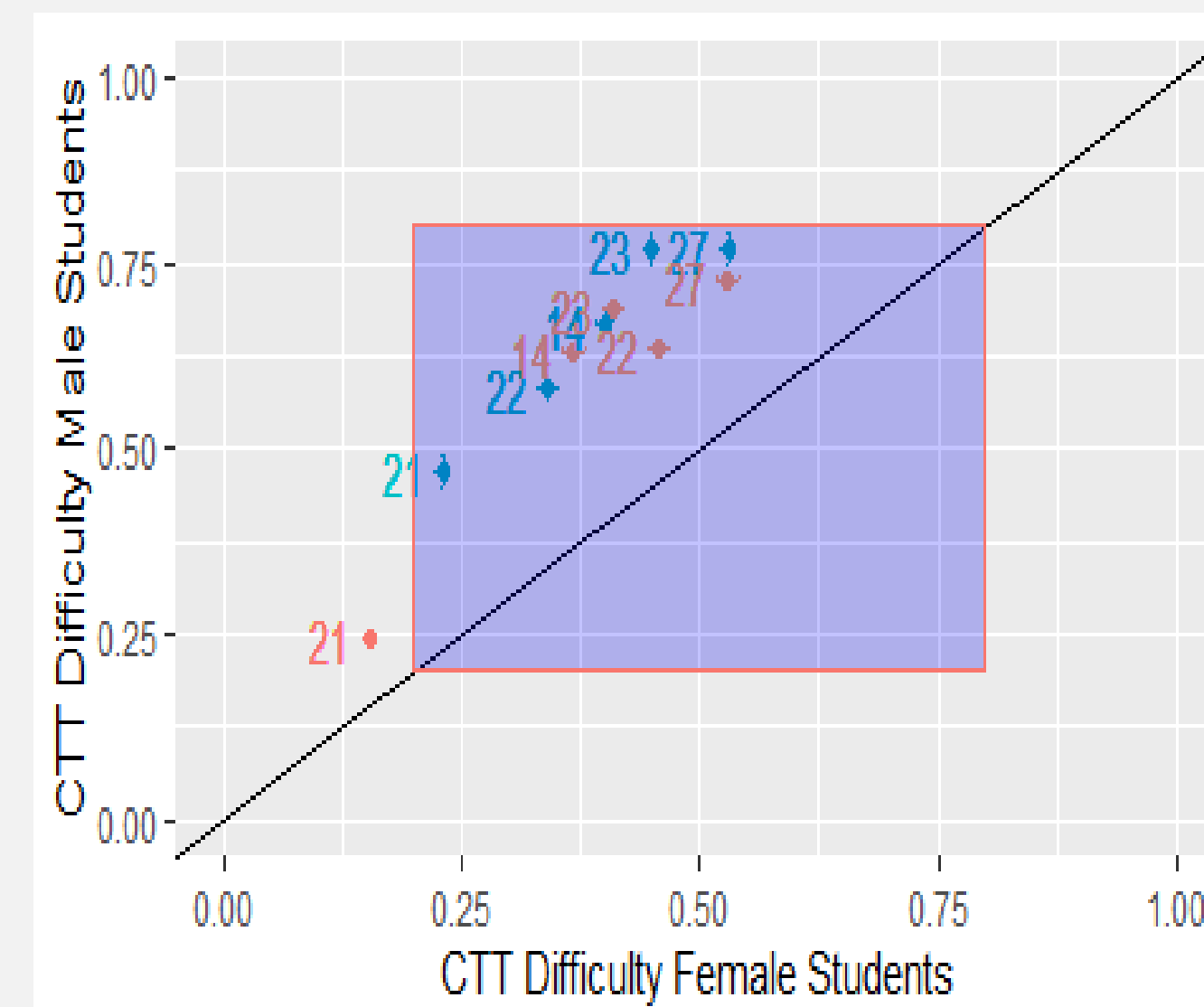
Gender	Pre or Post	Problematic Items
Female	Pre	5, 11, 13, 15, 17, 18, 21, 25, 26, 28, 30
	Post	1, 3, 6, 7, 8, 9, 10, 12, 16, 19, 21, 24, 25, 29
Male	Pre	5, 6, 17, 18, 21, 25, 26
	Post	1, 3, 6, 7, 8, 9, 10, 12, 13, 16, 19, 21, 24, 25, 29
Overall	Pre	5, 11, 17, 18, 21, 25, 26
	Post	1, 3, 6, 7, 8, 9, 10, 12, 13, 16, 19, 21, 24, 25, 29

Problematic items identified via CTT. Items highlighted in red are the problematic items found in common. Bold items are problematic for women, but do not show up overall.

- LASSO Dataset
- Traxler's Dataset
- Common Dataset

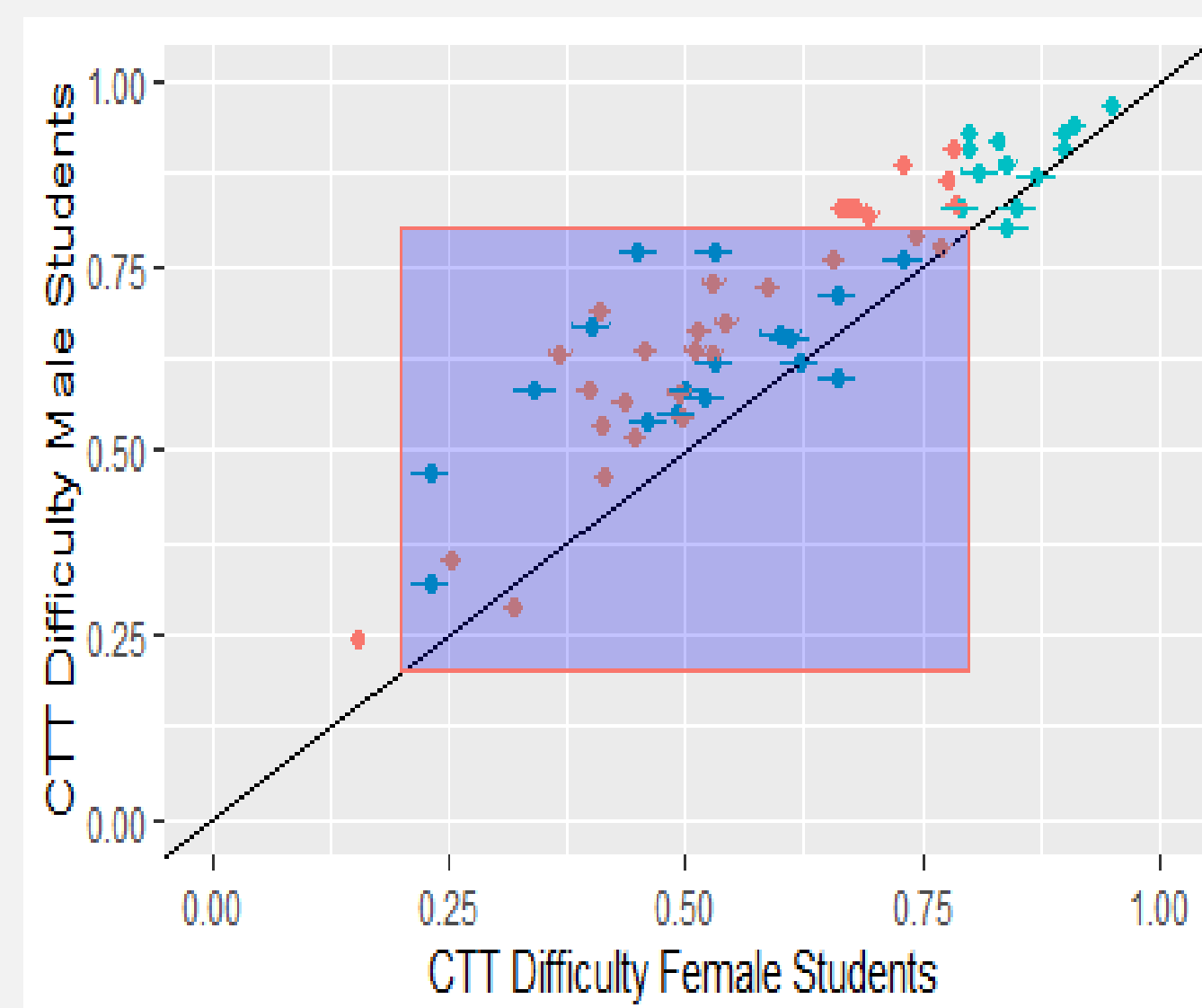
Most Items on the FCI Advantage Men

The 5 most problematic items found (via CTT) by Traxler were similarly problematic in LASSO. Higher CTT difficulty means an easier item. Fair items are inside the box.

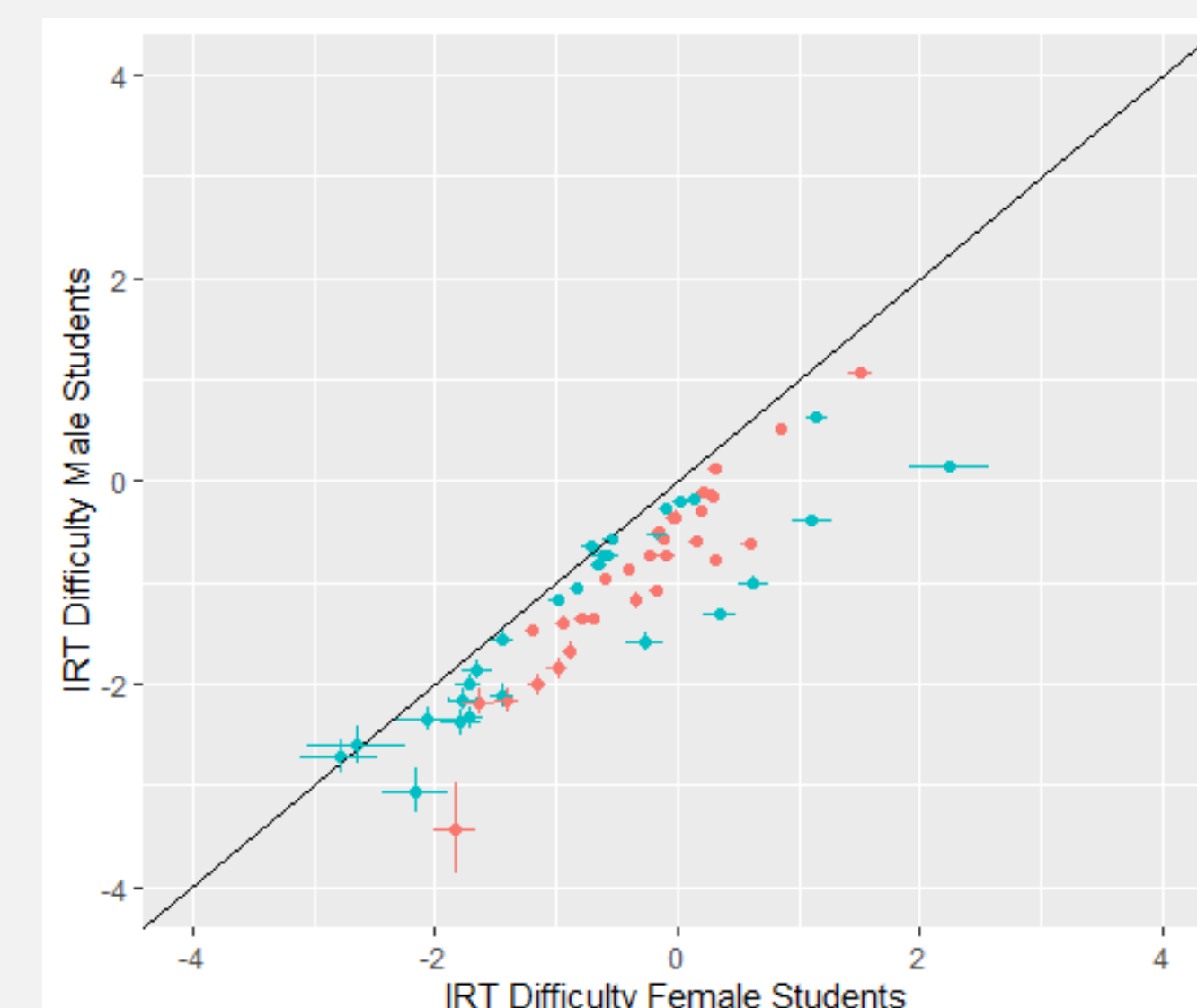


Traxler's Dataset

LASSO Dataset



Both sets of data, with all FCI questions. Fair items are inside the box. This is the CTT analysis. Higher CTT difficulty means an easier item.



Both sets of data, with all FCI questions. This is the IRT analysis. Lower IRT difficulty means an easier item.

Discussion

- Several items were found to be problematic between both studies.
 - Helps us verify that items Traxler *et al.* found to be problematic are problematic even with different institutions and collection methods.
- Items 14, 22, 23, and 27 were the most advantageous to men on the posttest in both studies.
- Both studies showed that there was a bias towards men for most items.
- Differences in results do not invalidate Traxler's findings.
- Similarities in results help validate Traxler's findings since there were many differences in the datasets.

Conclusions

- 4 of the 5 most problematic items found by Traxler were the most problematic in the LASSO dataset.
- Nearly all items on the FCI advantage men.
- We agree with Traxler that there are issues with the structure of the FCI that cause it to have negative effects on gender performance gaps.

References

- Traxler, R. Henderson, J. Stewart, G. Stewart, A. Papak, and R. Lindell, *Gender fairness within the Force Concept Inventory*, Phys. Rev. Phys. Educ. Res. 14, 010103 (2018).
- J. M. Nissen, M. Jariwala, E. W. Close, and B. Van Dusen, *Participation and performance on paper- and computer-based low-stakes assessments*, Int. J. STEM Educ. 5, 21 (2018). <https://learningassistantalliance.org/>

Acknowledgements

Material based on work supported by NSF DUE 1560142 and DUE 1852045. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF. A special thanks to the CIDER REU mentors and cohort!

