ITEM RESPONSE THEORY AND RANKING OF CONGRESSIONAL VOTING BEHAVIOR

HOLMES FINCH

GREGORY MARCHANT

ALICIA HAZELWOOD

BALL STATE UNIVERSITY

# THE USE OF RANKINGS IN SOCIAL POLICY ANALYSIS

- The ranking of individuals, states, and nations is a very common practice in reporting all manner of social policy data.

- For example, the U.S. government ranks states based upon their schools' performances on the tests associated with the National Assessment of Educational Progress (NAEP).

- Likewise, countries are ranked by their performance on educational measures (e.g., PISA), health care systems, economic output, happiness, and quality of democracy.

BALL STATE
UNIVERSITY

QUALITY OF RANKINGS

- An important issue to consider when creating or using such rankings, is the manner in which they are obtained.

- In some cases, the answer to this question is fairly straightforward, as it is based upon a single value, such as the score on an educational achievement measure.

- In other cases however, such is not the case, and rankings are based upon a much more opaque process involving the combining of several measures into a single index that is subsequently used to rank individuals.

- The goal of the current study is to investigate how Item Response Theory (IRT) might prove useful in such endeavors.

CHARACTERIZATION OF CONGRESSIONAL VOTING BEHAVIOR

- Political action committees, lobbying organizations, and academics routinely rate and rank members of the United State Congress in terms of their voting behavior.

- Results of these rating and ranking efforts are used to make decisions regarding the allocation of campaign donations, advertising efforts on behalf of (or in opposition to) specific candidates, and in targeting lobbying efforts.
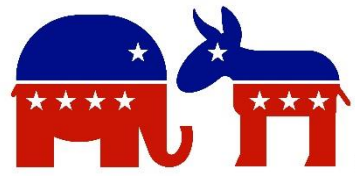
CHARACTERIZATION OF CONGRESSIONAL VOTING BEHAVIOR

- Frequently, members of congress are ranked based upon a simple summation of their votes for bills focused on selected issues (e.g., health care).

- In other instances an attempt is made to weight votes so those considered more important by the rating organization (e.g., the National Rifle Association) have a greater influence on the final score

- It is important to consider the methodology used to derive the weights.

- IRT might prove useful in providing an empirically defensible method for ranking legislative voting behavior, as well as for gaining deeper insights into the voting behavior itself.

GOALS OF THIS STUDY

1. Characterize congressional voting patterns using IRT.

2. Investigate anomalous voting behavior by political party, using differential item functioning (DIF).

3. Identify legislators who vote in anomalous ways.

4. Compare IRT-based ranking of congressional voting behavior with that produced by a political action committee.

METHODOLOGY: SAMPLE AND DATA

- Data were taken from records of 15 votes by 434 members of the 114th U.S. Congress (2016).

- The votes, identified by the conservative political advocacy organization Freedom Works (FW), were scored as 1 (yes) and 0 (no).

- FW calculated a total score based upon these 15 votes with higher scores reflecting a more conservative voting record.

- Members of congress were then ranked based upon these scores.

METHODOLOGY: DATA ANALYSIS

- The first step of the analysis involved determining the appropriate IRT model to fit to the data.

- Next, uniform DIF based upon political party was investigated using the ETS DIF classification heuristic based upon the Mantel-Haenszel (MH) procedure, with scale purification.

- Third, anomalous voting patterns for individual legislators were investigated using the l_z^* person fit statistic.

- Finally, based upon the latent variable estimated by the IRT model, the legislators were ranked according to their level of conservatism.

BALL STATE UNIVERSITY

RESULTS: IRT MODEL

- The Rasch and 2-parameter logistic (2PL) models were each fit to the data, and the optimal one was selected based upon the AIC and BIC values.

Model	AIC	BIC
Rasch	5653.97	5705.63
2PL	5715.00	5827.00

- Given these results, the Rasch model was identified as the better fitting of the two.

- The bootstrapped Chi-square goodness of fit test was not statistically significant (a=0.05), suggesting that the Rasch model adequately fits the data.

BALL STATE
UNIVERSITY

RESULTS: IRT MODEL

| Bill	Overall
Override Veto of Healthcare Freedom	-5.87 (0.59)
Common Sense Nutrition	-0.39 (0.13)
Gosar Amendment	-0.34 (0.13)
Restore Healthcare Freedom	-0.42 (0.13)
Sandford Amendment	2.59 (0.17)
Perry Amendment	0.35 (0.13)
Palmer Amendment	2.00 (0.15)
Email Privacy Act	0.20 (0.13)
Preventing IRS Abuse	0.81 (0.17)
Massie Amendment	-0.41 (0.13)
Buck Amendment	0.65 (0.13)
Duffy Amendment	0.21 (0.13)
Anti-Terrorism Info Sharing	0.36 (0.13)

1. Override veto of healthcare Freedom Act – Override veto of bill repealing Obamacare.

2. Sandford amendment - Provide stipend for military personnel to buy footwear of choice, whether or not it was made in USA.

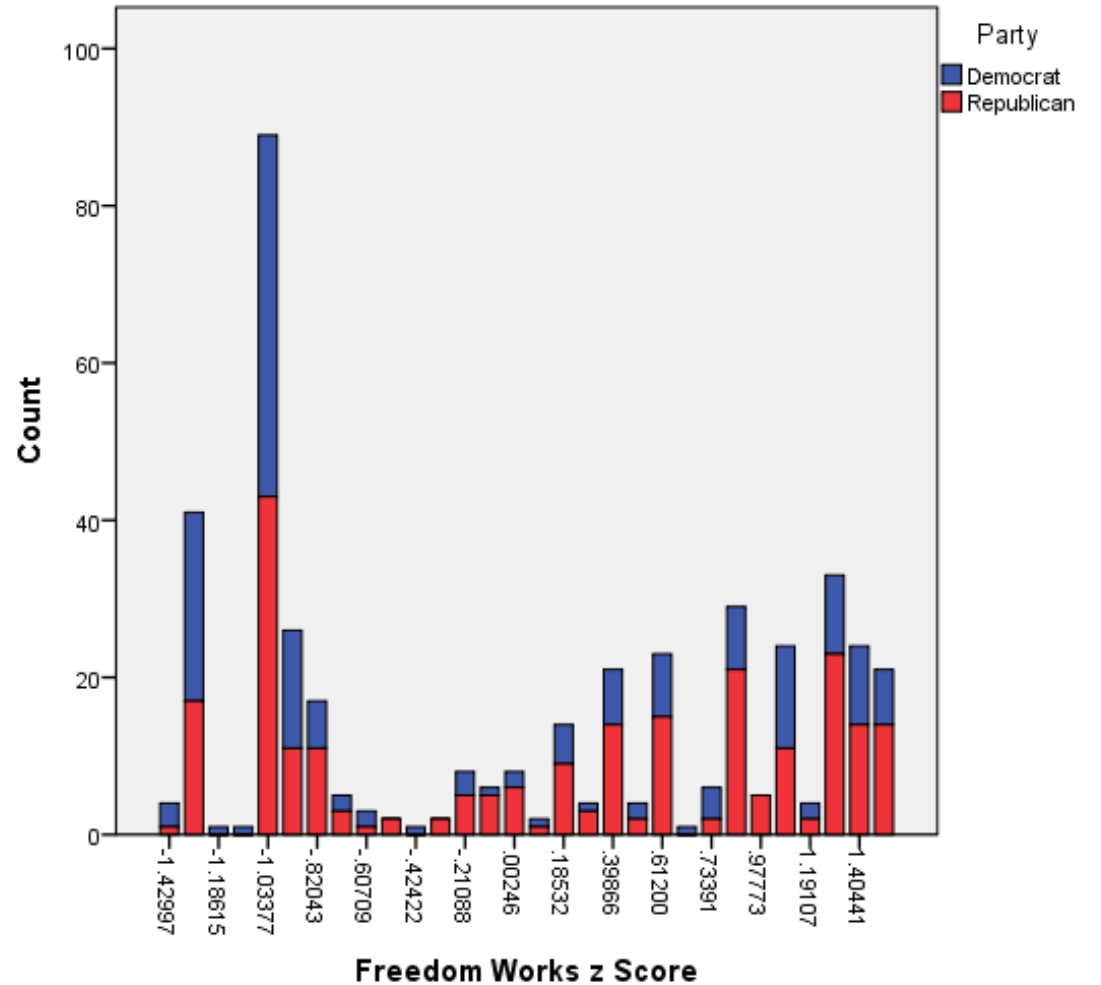3. Palmer amendment - Restricts EPA use of funds to engage in criminal enforcement of environmental laws.
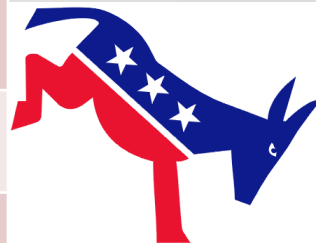
RESULTS: CONSERVATISM DISTRIBUTION

Distribution of IRT Conservatism Trait
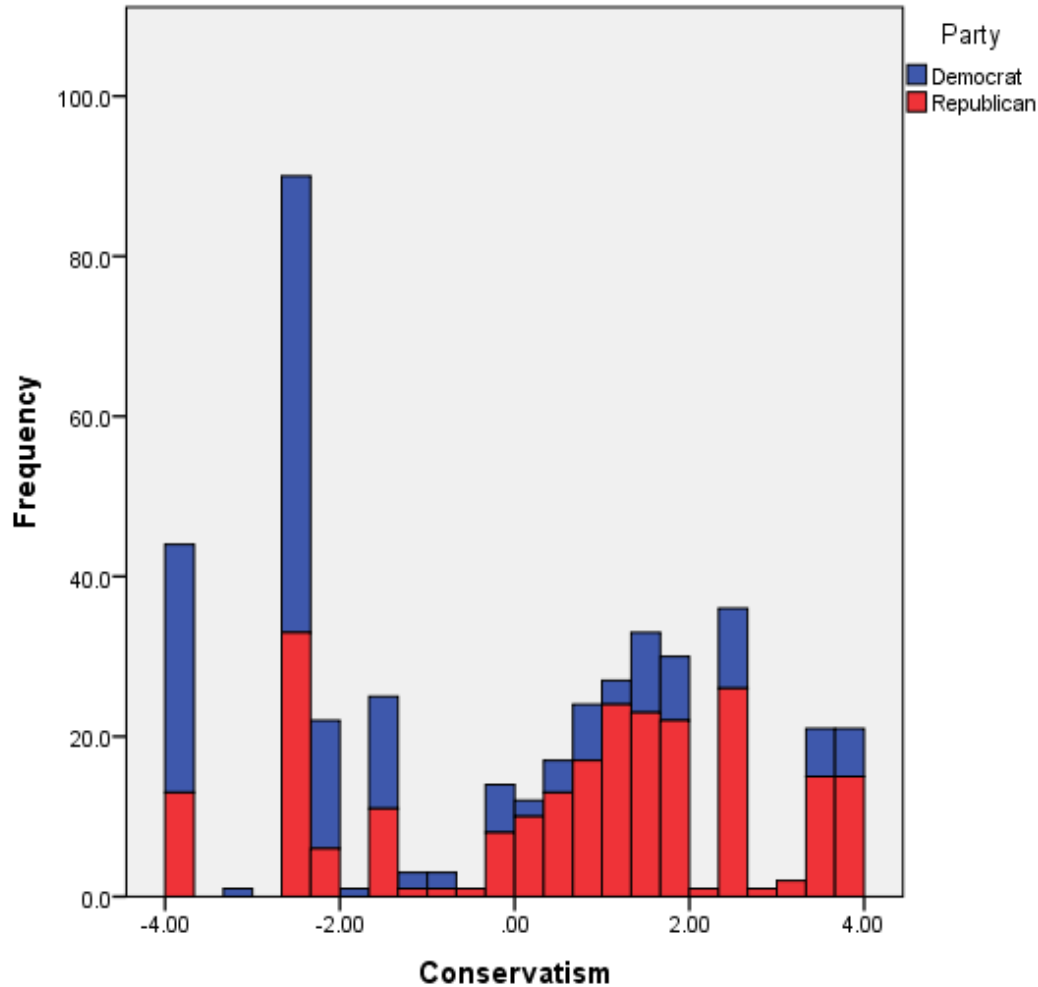
Distribution of FW z Scores

RESULTS: CONSERVATISM DISTRIBUTION

- Friedman's test retained the null hypothesis that the distributions of the two variables were the same.

- This meant that the FW score and IRT conservatism latent trait share a common distributional form.

- Conversely, the Kolmogorov-Smirnov test revealed that the distributions of the FW score and the IRT conservatism trait were significantly different from one another with respect to political party.

- In other words, for each measure of conservatism, Republicans and Democrats have different latent trait distributions for the scores.

RESULTS: CONSERVATISM DISTRIBUTION
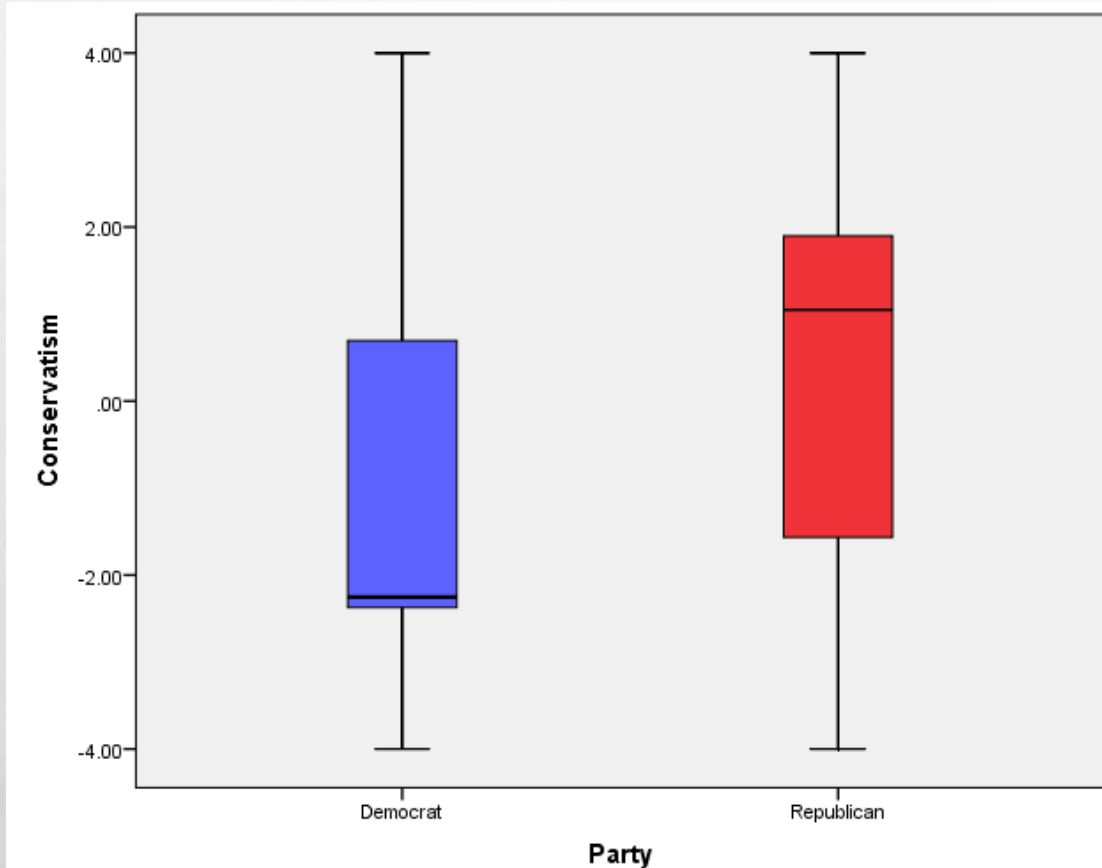
- CONFIDENCE INTERVALS FOR MEAN CONSERVATISM:
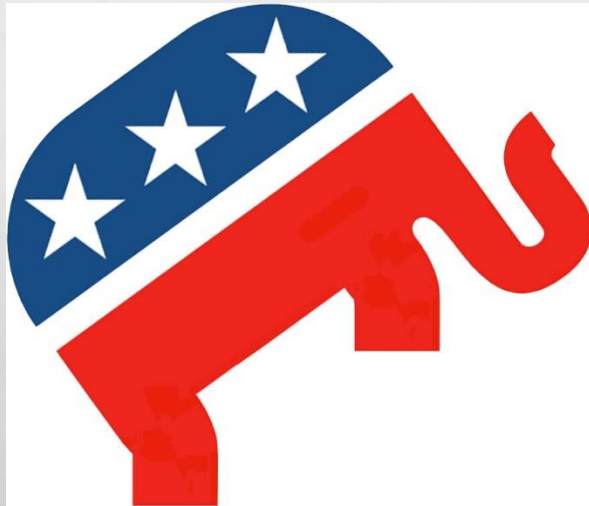
  - DEMOCRATS: -1.505, -0.841
  - REPUBLICANS: 0.301, 0.850

- MEDIAN CONSERVATISM:

  - DEMOCRATS: -2.255
  - REPUBLICANS: 1.048

RESULTS: DIF

Bill	Democratic	Republican	ETS Class
Override Veto of Healthcare Freedom	-4.84 (0.60)	-16.21 (12.70)	C
Common Sense Nutrition	0.85 (0.21)	-1.33 (0.19)	C
Gosar Amendment	0.91 (0.21)	-1.26 (0.19)	C
Restore Healthcare Freedom	0.82 (0.21)	-1.36 (0.19)	B
Sandford Amendment	3.62 (0.34)	2.02 (0.20)	B
Perry Amendment	1.39 (0.22)	-0.37 (0.17)	B
Palmer Amendment	2.45 (0.26)	1.66 (0.18)	B
Email Privacy Act	-0.40 (0.19)	0.56 (0.17)	A
Preventing IRS Abuse	1.10 (0.21)	0.57 (0.17)	A
Massie Amendment	0.86 (0.21)	-1.37 (0.19)	A
Buck Amendment	1.56 (0.23)	0.05 (0.17)	A
Duffy Amendment	1.46 (0.23)	-0.62 (0.17)	A
Anti-Terrorism Info Sharing	1.57 (0.23)	-0.46 (0.17)	A

RESULTS: DIF

LARGE DIF VOTES

1. Override veto of Healthcare Freedom Act
2. Common sense nutrition act – Relax FDA regulation compelling restaurants to provide calorie information.
3. Gosar amendment – Prohibit paying performance bonuses to IRS employees.

MEDIUM DIF VOTES

1. Restore Healthcare Freedom Act
2. Sandford amendment
3. Perry amendment – Reduce EPA appropriations by 17%.
4. Palmer amendment

RESULTS: PERSON FIT

- Anomalous voting patterns were identified using the I_z^* statistic.

- Values of -1.65 or less indicated that the Rasch model provided poor fit to the voting behavior of specific legislators.

- Results revealed that the model provided poor fit for 19 members of congress, 13 of whom were Democrats.

RESULTS: PERSON FIT

Distribution of Iz* Scores
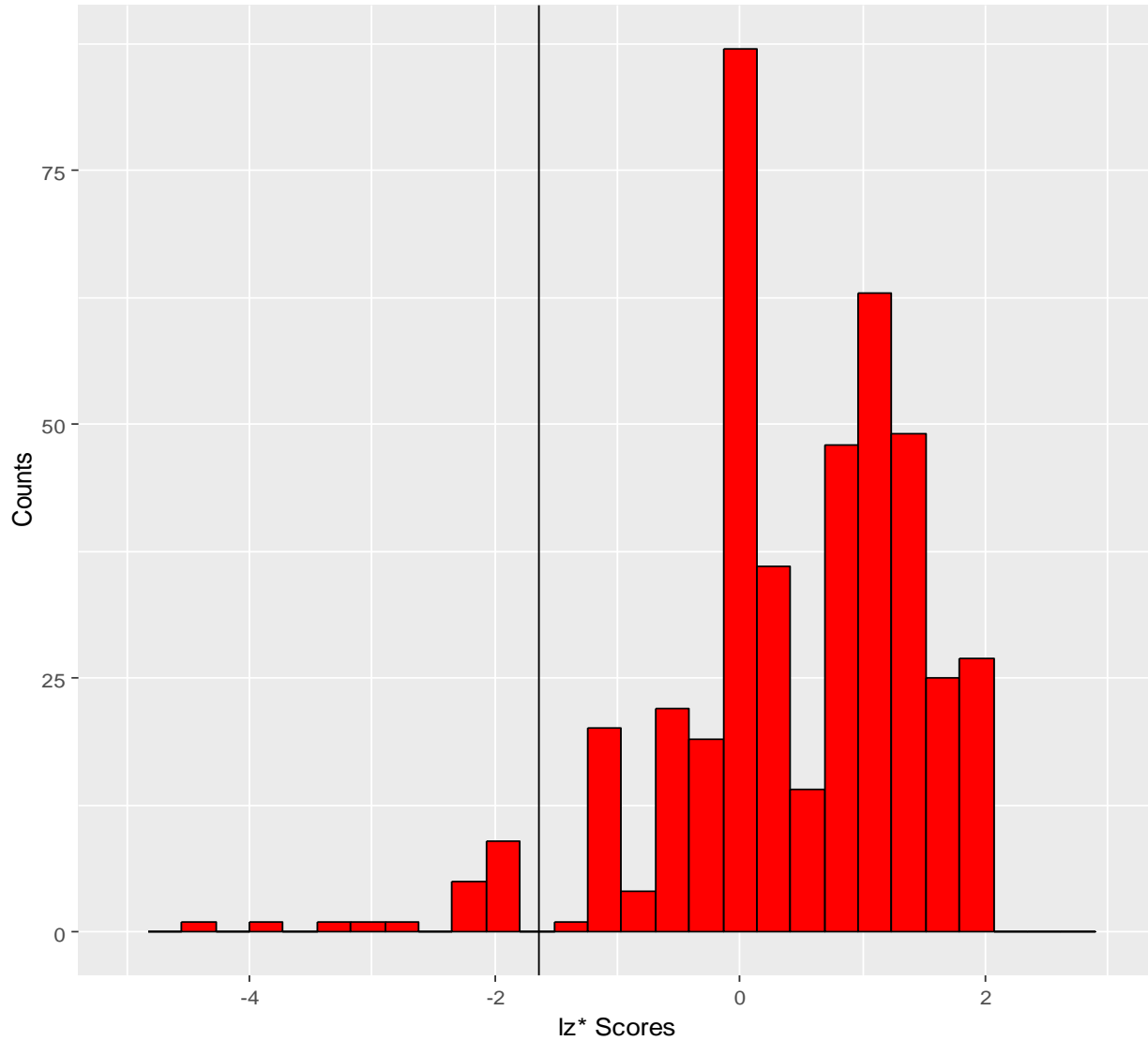
| Name	Party	State	Race	Gender	IRT Conservatism
Stephen Finchers	R	TN	White	Male	3.22
Steve King	R	IA	White	Male	1.85
Gerry Connolly	D	VA	White	Male	-0.91
Earl Blumenauer	D	OR	White	Male	-1.08
Lloyd Doggett	D	TX	White	Male	-1.45
Randy Weber	R	TX	White	Male	-1.45
Dan Kildee	D	MI	White	Male	-1.60
Jim McGovern	D	MA	White	Male	-1.60
Diana DeGette	D	CO	White	Female	-1.60
Renee Ellmers	R	NC	White	Female	-1.60
Brenda Lawrence	D	MI	African American	Female	-1.61
Lacy Clay	D	MO	African American	Male	-2.37
Adam Schiff	D	CA	White	Male	-4.00
Kurt Schrader	D	OR	White	Male	-4.00
J. Sensenbreener	R	WI	White	Male	-4.00
Steve Womack	R	AK	White	Male	-4.00

RESULTS: RANKING OF LEGISLATORS BASED ON VOTING BEHAVIOR

- The final goal of this research was to use conservatism scores (either FW or IRT based) in order to rank members of congress.

- The correlation between the FW score and the Rasch latent trait estimate was 0.34.

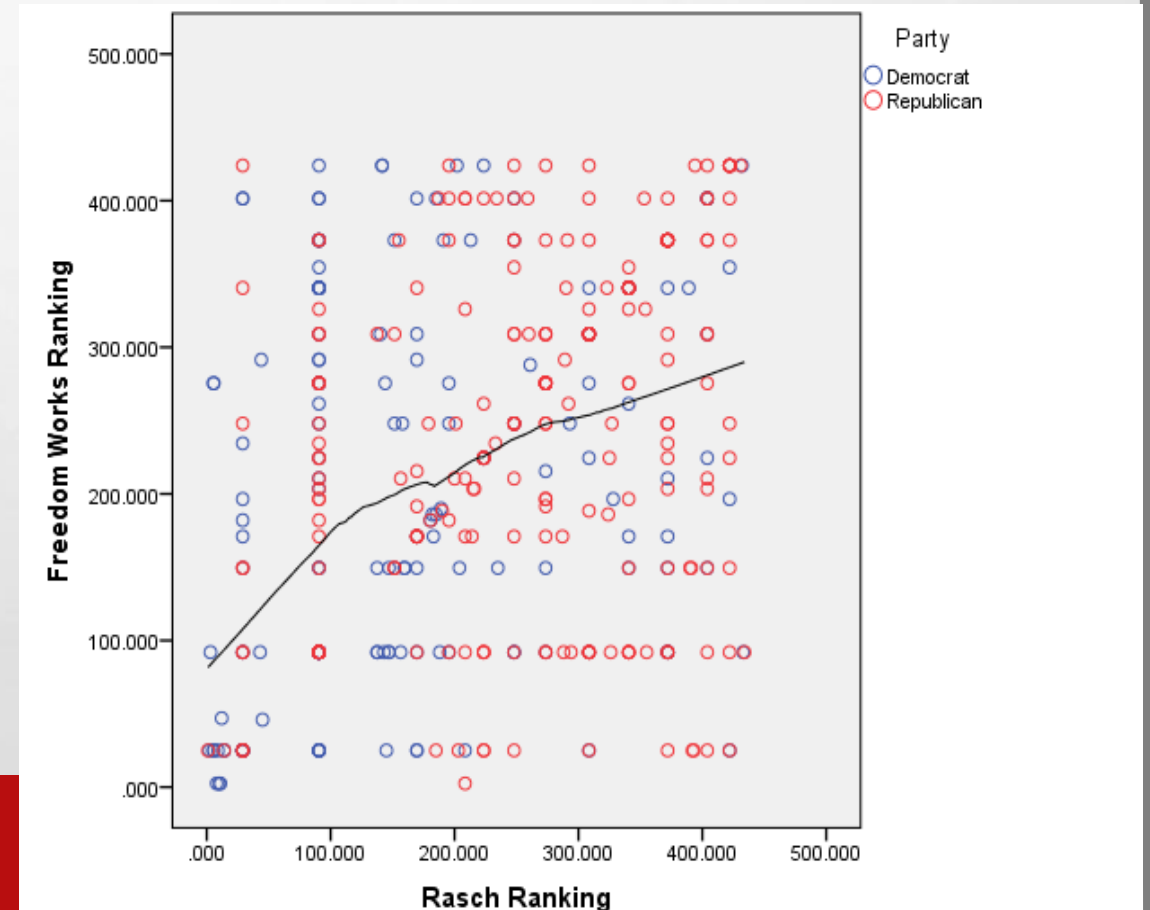- The correlation between the FW and Rasch based rankings was 0.35.

RESULTS: RANKING OF LEGISLATORS BASED ON VOTING BEHAVIOR
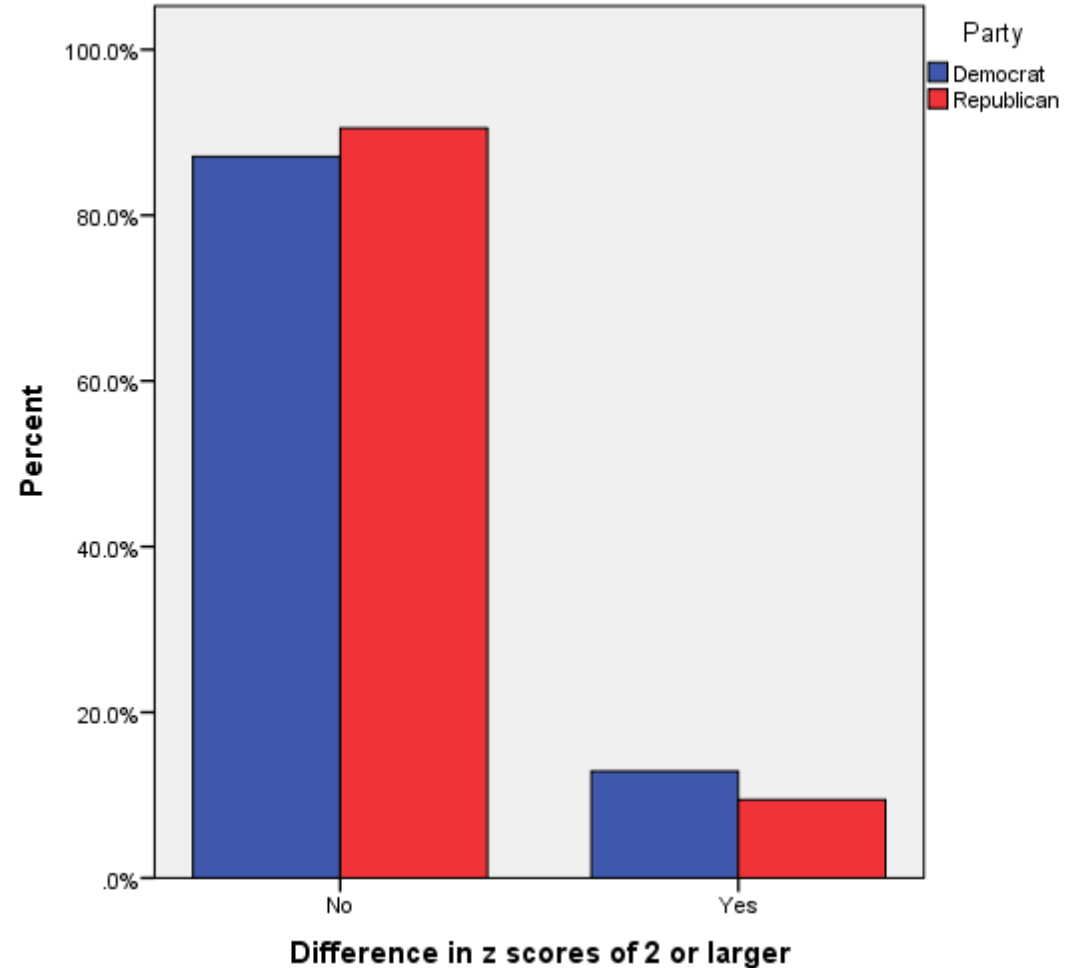
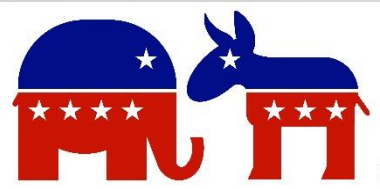## FREEDOM WORKS SCORE BY RASCH CONSERVATISM LATENT TRAIT
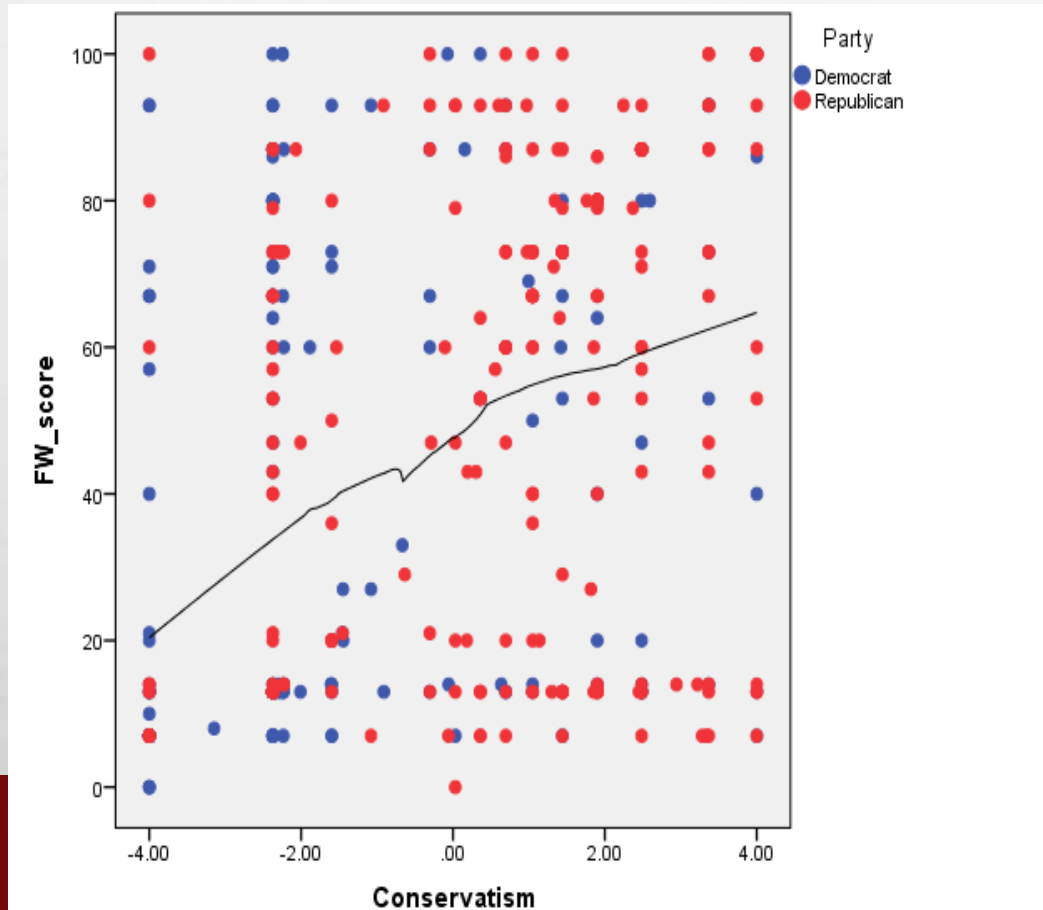
FREEDOM WORKS RANKING BY RASCH CONSERVATISM RANKING

RESULTS: DIFFERENCE IN FW AND IRT SCORES BY PARTY

- The FW scores were converted to z scores and were then directly compared to the IRT latent trait estimates.

- A total of 46 members of congress had absolute differences in z scores of 2 or more.

RESULTS: DIFFERENCE IN FW AND IRT SCORES BY PARTY

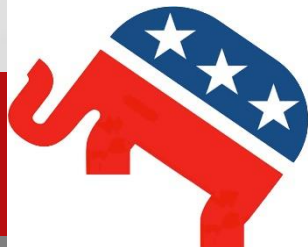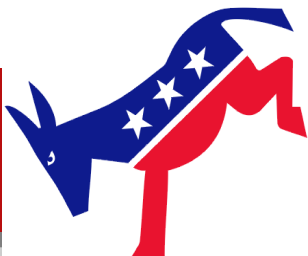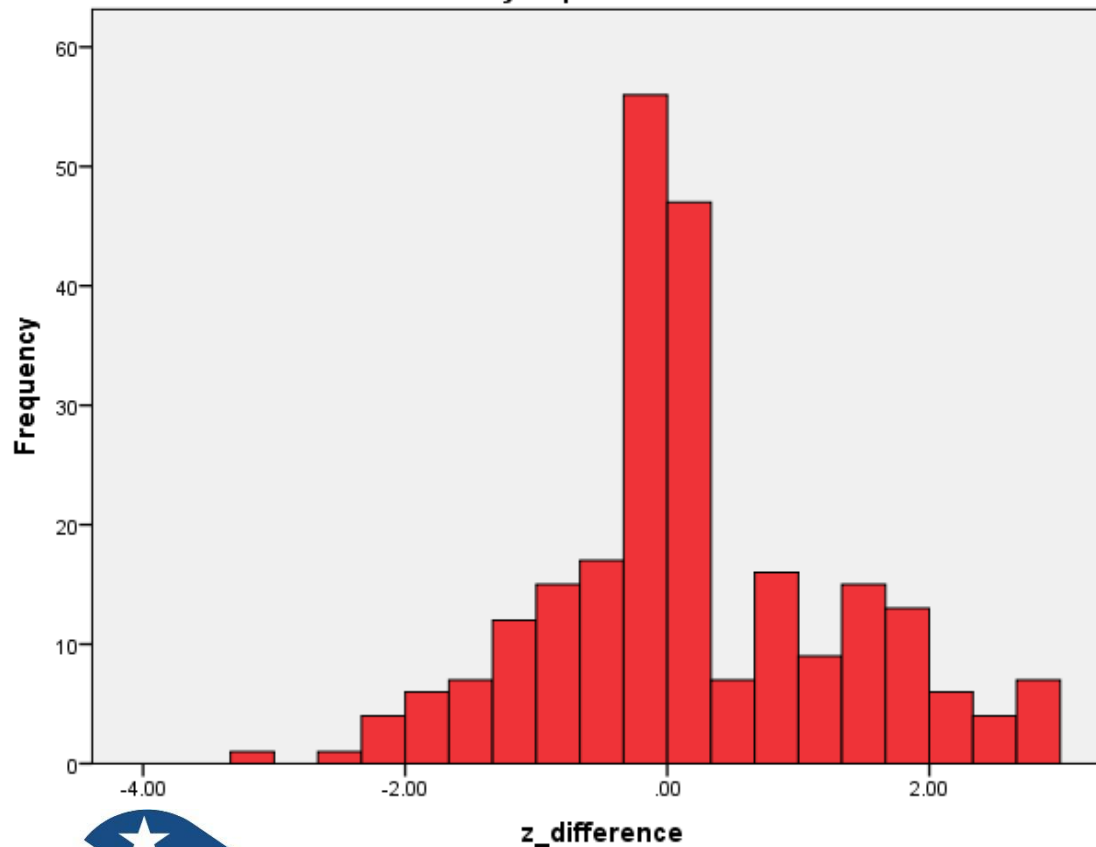- Results for the individuals with the largest difference between the IRT and FW scores.

Name	Party	FW score z	FW rank	IRT score	IRT rank	z Score Diff
Duffy	R	1.62	424	-1.61	29	-3.23
Langevin	D	1.40	402	-1.61	29	-3.02
Ruiz	D	1.40	402	-1.61	29	-3.02
Farr	D	-1.22	25	1.76	422	2.97
Griffith	R	-1.22	25	1.76	422	2.97

RESULTS: DIFFERENCE IN FW AND IRT SCORES BY PARTY
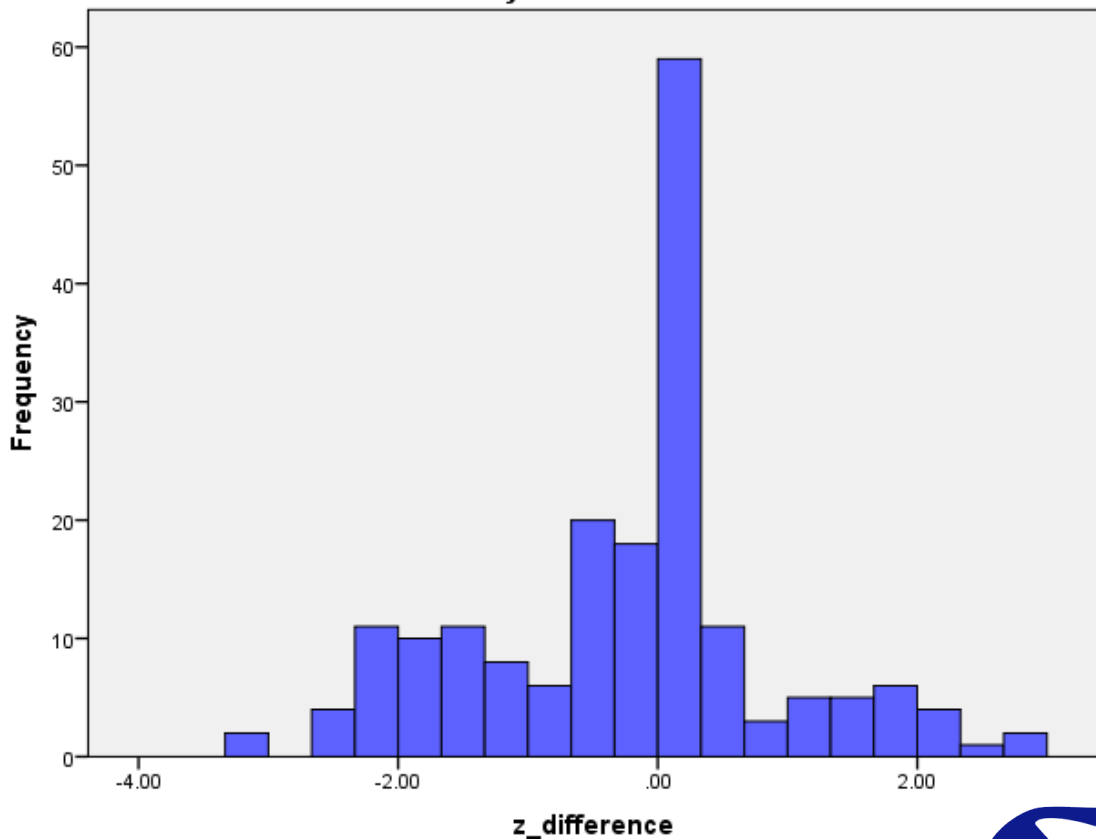
CONCLUSIONS

- IRT based modeling allowed for the estimation of the latent trait of interest (i.e., conservatism), as well as the identification of individual legislators for whom the model did not provide good fit, despite the fact that overall the model did fit well.

- Although the conservatism estimate for some of these individuals was unusually high or low, for others no obviously discernible cause for the misfit was evident.

- The relationship between the FW score and IRT-based conservatism trait estimate was of moderate size, though for some individuals there was a relatively large difference between the two scores.

CONCLUSIONS

- In addition to yielding estimates of the underlying latent trait, IRT also provides tools for identification of specific votes that reflect the trait of interest (e.g., conservatism) in anomalous ways for members of different political parties.

- Several such votes were found, and in each case they were more difficult for Democrats to support, when members of the two parties were matched on their underlying conservatism.

CONCLUSIONS

- The overarching purpose of this research was to demonstrate the utility of IRT based tools in the study of legislative voting behavior.

- IRT is a powerful tool for estimating the latent trait of interest (e.g., conservatism), which can then be used to rank individuals, or otherwise make direct comparisons among them.

- These rankings may better reflect the relative positioning of legislators than do more ad hoc methods of scoring their voting behaviors.

conservative base