

**Can a Non-Probability Sample Ever  
be Useful for Representing a  
Population? Comparing Probability  
and Non-Probability Samples of  
Recent College Graduates.**

# Project Team

## Heldrich Ctr. Rutgers

- Cliff Zukin
- Carl Van Horn
- Jessica Godofsky

## Knowledge Networks

- Mike Dennis
- Wendy Mansfield
- Sergei Rodkin

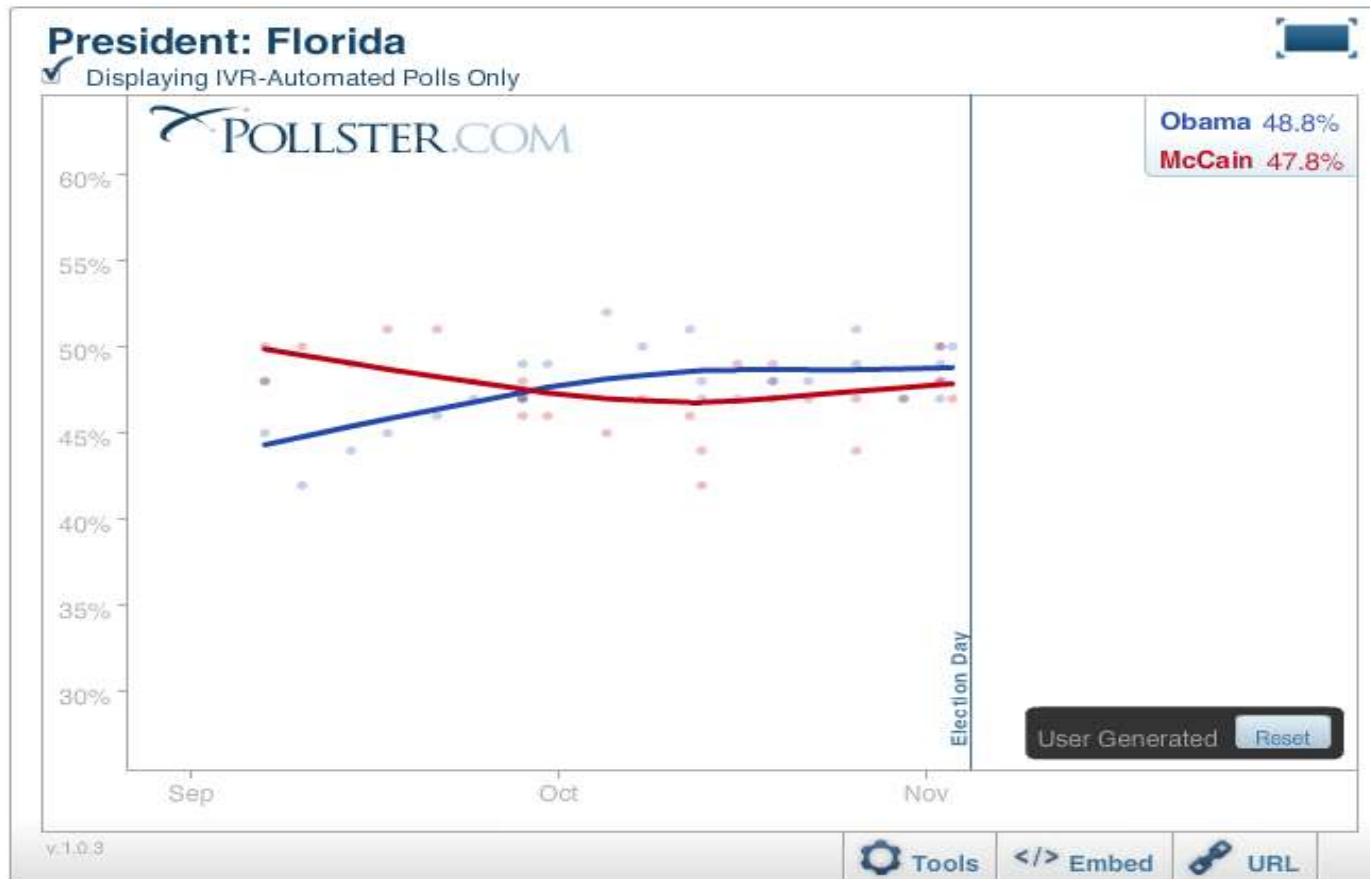
# Genesis

- **A Research Problem:**
- “How are Recent College Graduates (RCG) faring in the Great Recession, and in general?”
  
- **A Research Question:**
- Why do surveys (continue to) work with a response rate of 10 to 15%?

# Evidence that Surveys Still Work

- Pew Gold Standard Survey Experiments
  - In 77 out of 84 comparable items, the two surveys yielded results that were statistically “indistinguishable” Keeter 2006
- Time Series Measurement
  - Generally stable
- Election Polling
  - One estimate is as good as the other

# IVR Polls Today's Old Women

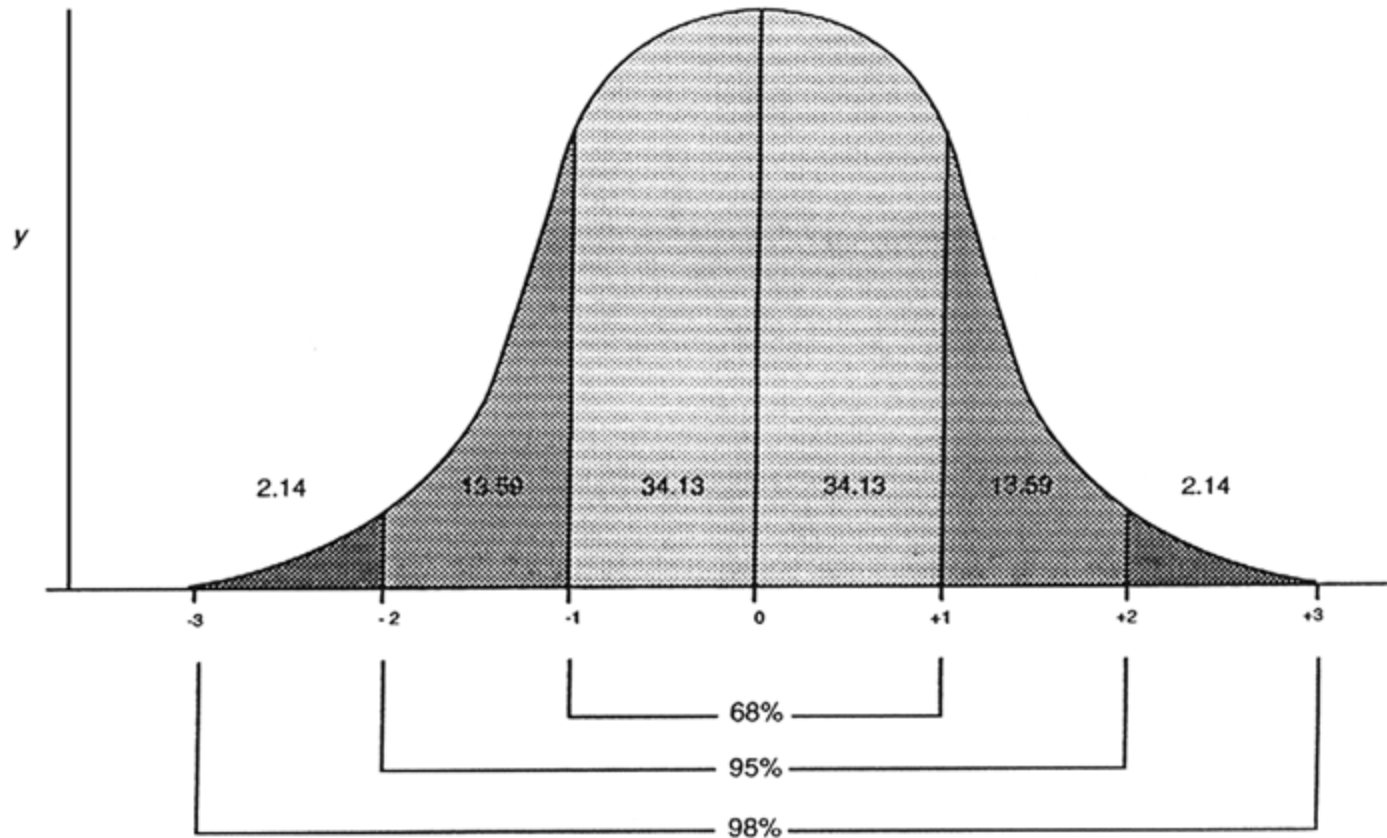


# Live Interviewer, multiple day



# Why Surveys Work--Cliff's Notes

Probability theory, large numbers, good response rate



# The Notion of Sampling

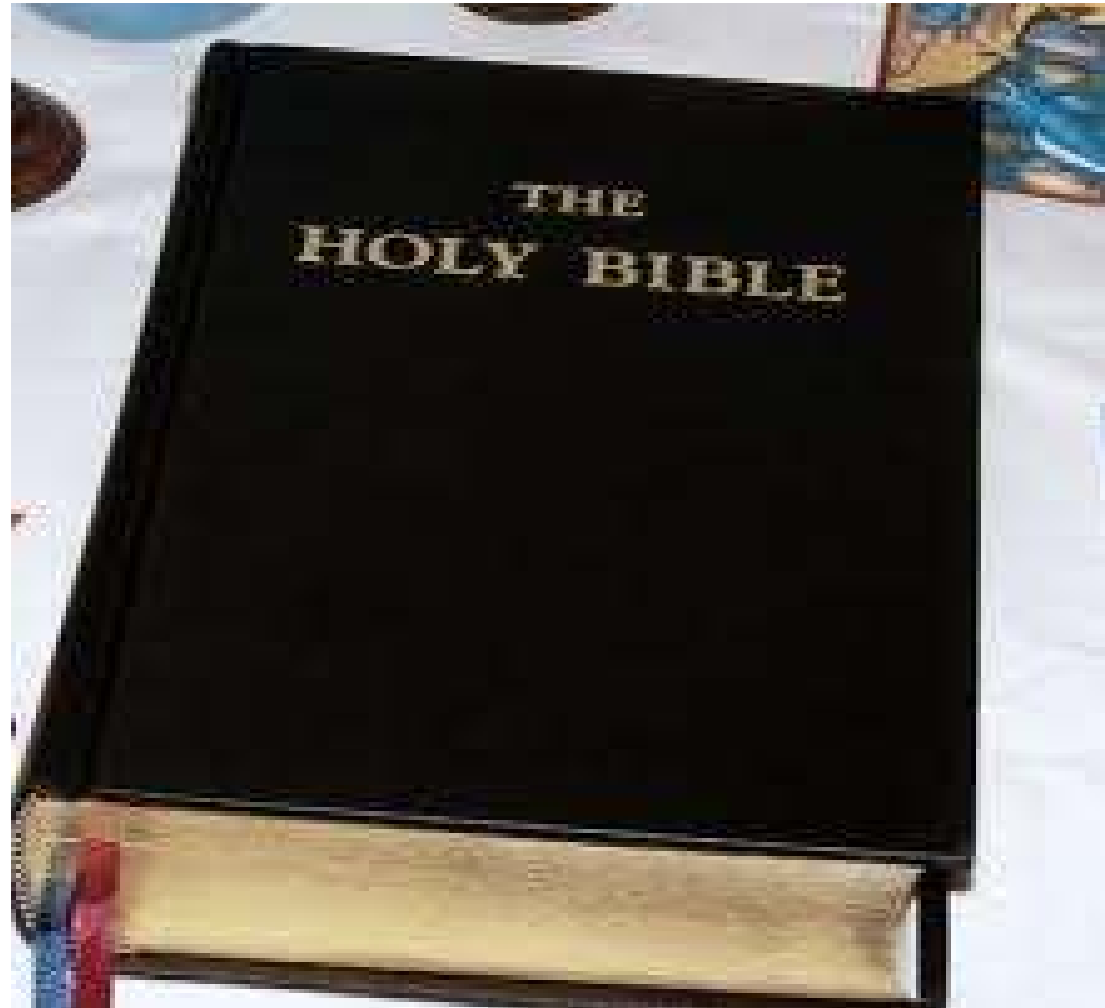
- Bloody Winers
- Polls are estimates
- A sample is drawn to represent an underlying population
- The sample must be **representative**

# Good and Bad Samples

- Probability Samples are GOOD
  - Scientific—known chance of inclusion
  - Generalizable
  - Random, or thereabouts, selection
- Non-Probability Samples are BAD
  - Not Representative
  - Self-Selection
  - People I talked to said...

# Background: Pasek and Krosnick (2010)

- Compared online, non-prob sample with telephone prob sample on the public's willingness to complete 2010 Census.  
Concerns included:
  - Representativeness
  - Differences in attitudinal measures/population values: proportions in most common responses, relationships between variables, changes in longitudinal variables



# Dillman et al: V3, Verse 343: $\pm 1$

- “In summary, volunteer Internet panels are developed around a very different model from the four-cornerstone error model followed in this book.”
- Coverage Error
- Sampling Error
- Measurement Error
- Non-response Error

# Properties & Principles

James Surowiecki *The Wisdom of Crowds*

## PROBABILITY SAMPLE

- Large n
- Diversity of Opinion or decentralization
- Independence of “members”
- Aggregation of views
- Known probability of being selected

## OPT IN SAMPLE

- Large n
- Diversity of Opinion or decentralization
- Independence of “members”
- Aggregation of views

# Methods

- Mode: Web survey
- Sample: Online panels of U.S. residents
- Sample Selection
  - Age 22 to 29
  - Bachelor's degree
- Screening Criteria
  - Bachelor's degree obtained from 2006 to 2010

# Sample

- Online probability-based panel (KnowledgePanel)
  - Recruited via ABS and RDD
  - Laptop and Internet-access provided to non-Internet households
- Online opt-in panel
  - Recruited via multi-sourced online approach (including banner ads, pop ups, and messages on web sites encouraging people to give their opinion, with a chance to win a prize)

# Data Collection

- Fielded March 15, 2011 – April 5, 2011

Stage	Probability Sample	Opt-In Sample
Sample	1,181	Not available
Screened	881	1,097
Cooperation rate	75%	Usually 5-7%
Surveyed	571	526
Eligibility rate	65%	48%

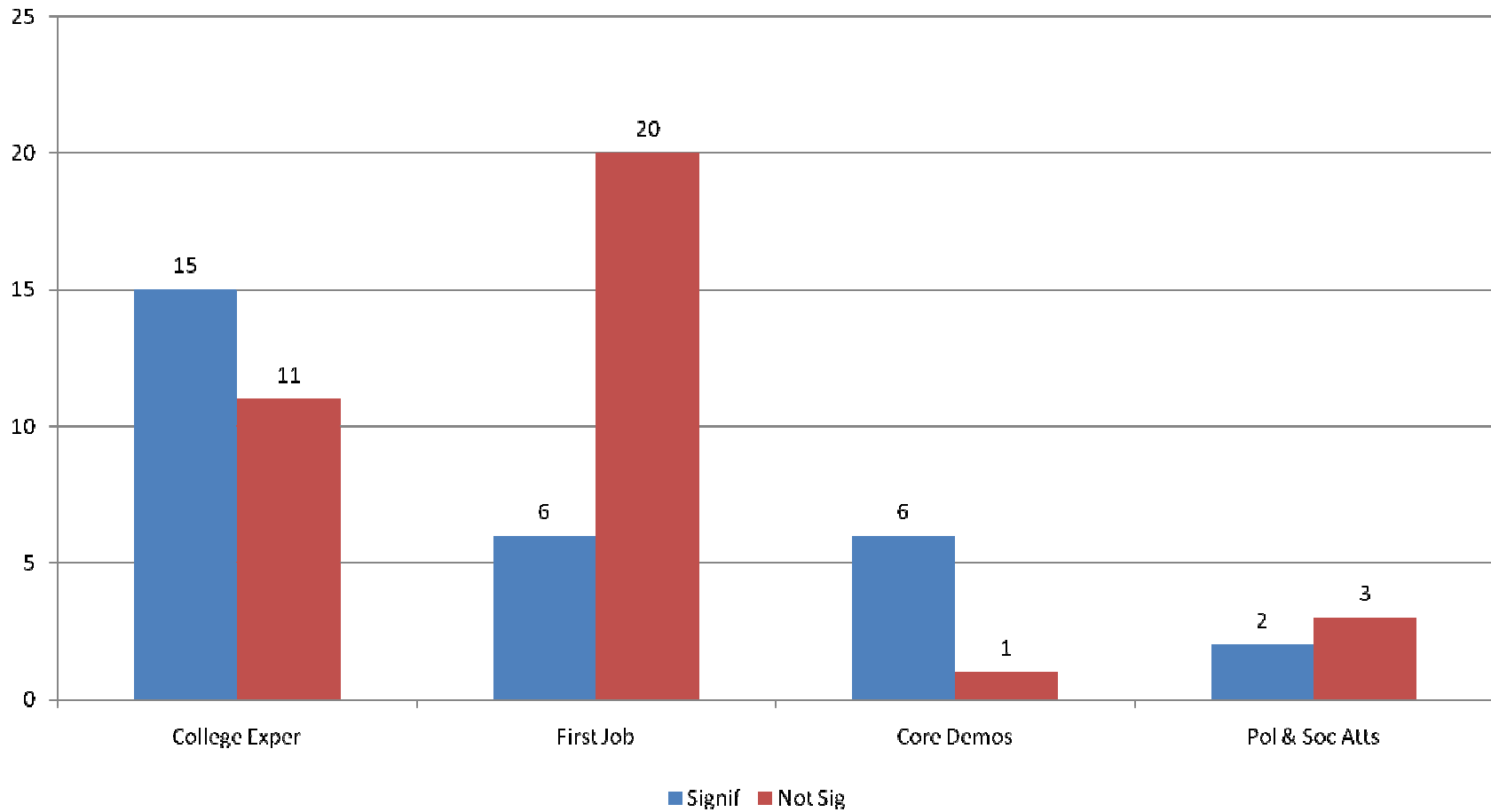
# Weighting

- Weighted probability sample using CPS benchmarks (gender by age, race/ethnicity, Census region, metro status)
- Weighted opt-in sample to look like weighted qualified probability respondents by controlling for gender, age, race/ethnicity, Census region, and metro status within the pre-recession and recession groups and on the overall level for graduation year, gender, age, race/ethnicity, Census region, and metro status.

# Modules on the Survey

- College Experiences (26)
- First Job Experiences (26)
- Current Job Experiences
- The Unemployed & NILF
- Core Personal Demographics (7)
- Political & Social Attitudes (5)

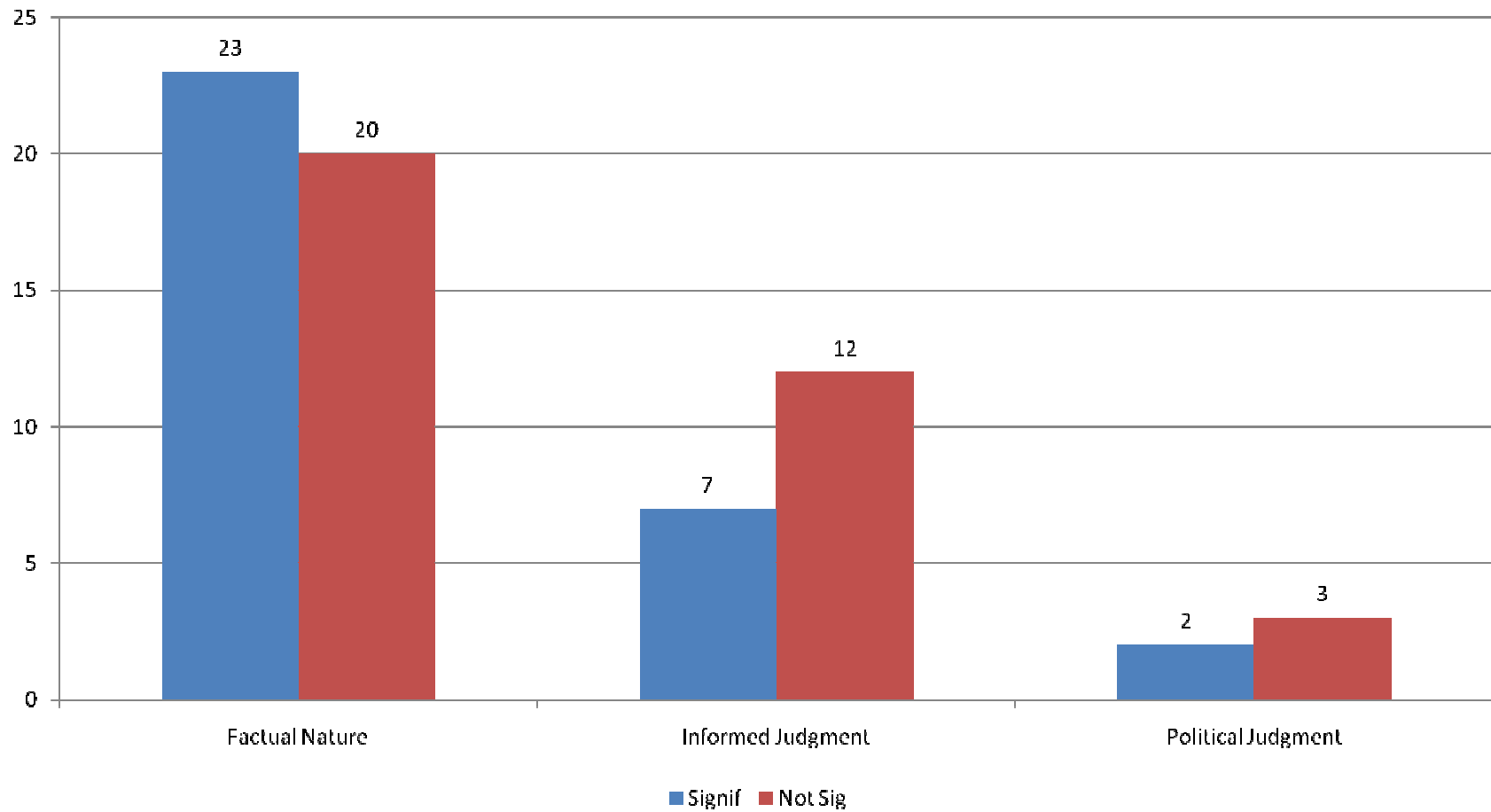
# Modules Chi<sup>2</sup> Differences



# Item Sort1 Chi <sup>2</sup>

- Factual (43)
- Informed Judgment (19)
- Political Judgment (5)

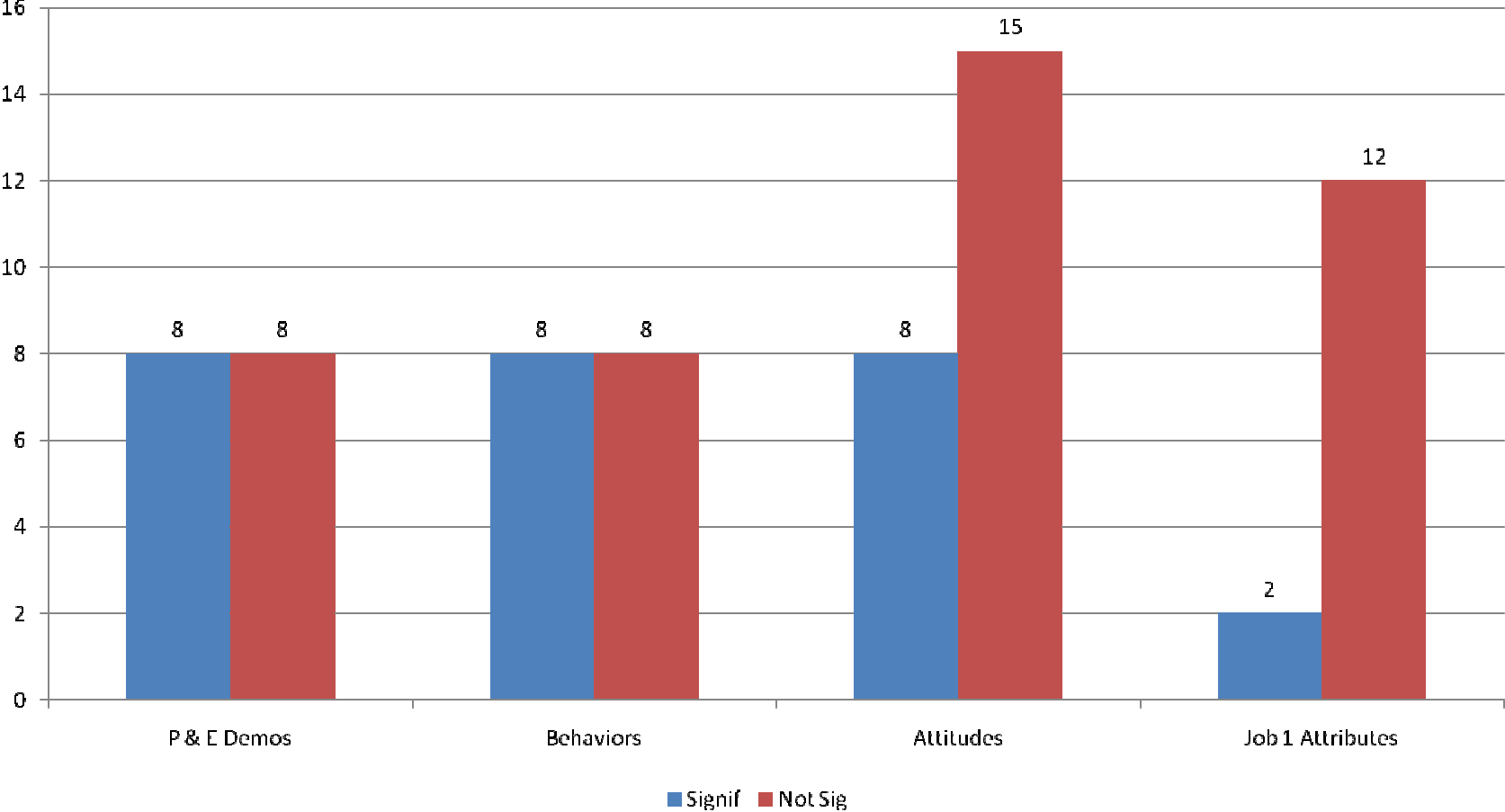
# Sort 1 Distribution



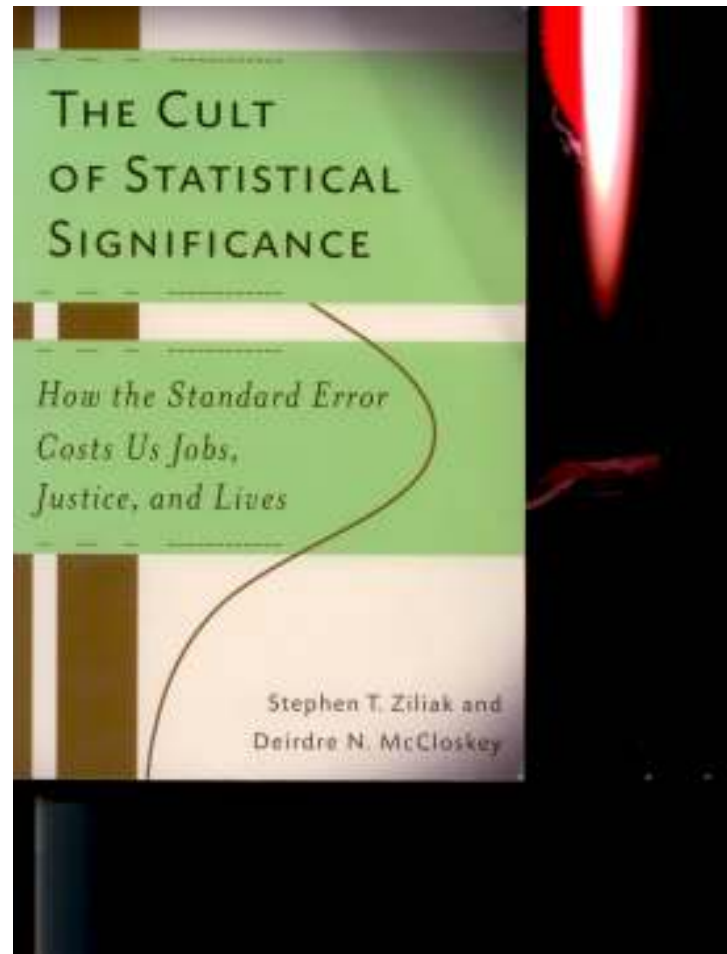
## Item Sort 2 Chi<sup>2</sup>

- Personal & Educational Demographics (16)
- Behaviors (16)
- Attitudes (23)
- First job attributes (14)

# Sort 2 Distribution



# Are the Differences Significant?



# Just 2 “Little” Problems: Employment

	Prob	NonProb	Prob	NonProb
	Wtd	Wtd	UnWtd	UnWtd
<b>Employed full-time</b>	<b>53</b>	<b>53</b>	<b>53</b>	<b>50</b>
Working PT want FT	7	13	7	13
Working PT, OK	5	4	5	4
Self-employed	3	3	3	3
More Ed, Working	14	3	14	3
More Ed, Not Working	7	4	8	4
Unemployed --looking	7	15	5	17
Unemployed -- not looking	2	6	3	6

# Just 2 “Little” Problems: Education

	Prob	NonProb	Prob	NonProb
	Wtd	Wtd	UnWtd	UnWtd
Attended Public College	65	72	66	73
Attend > 1 College	39	28	39	29
Attend community college	15	15	16	15
Worked FT while in school	23	23	22	23
Owed money at graduation	59	60	58	61
Took a loan (government)	36	44	38	45

# Background: AAPOR Task Force on Opt-In Online Panels (March 2010)

- Conclusions:
  - Non-prob online panels not for estimating population parameters
  - Mode effects explain some differences, but self-selection bias and undercoverage also have significant impacts
  - Weighting adjustments have varying effects on improving results