Item Order Effects in Student Surveys:
How Bad Are They and Is Randomization the Answer?

Corrie Potter

Stanford University

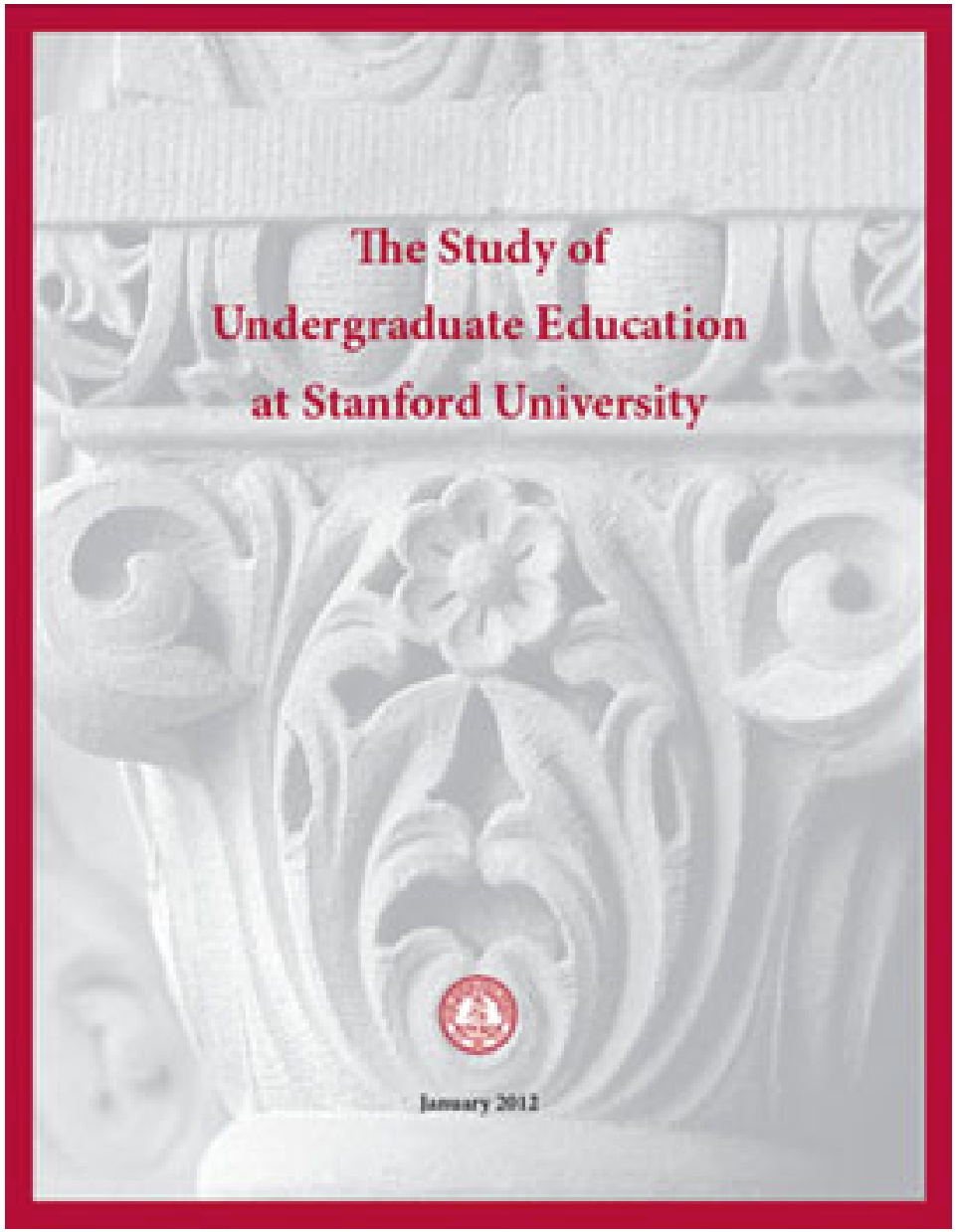
November 21, 2013

Overview

- Context
- Motivation
- Study Details
  - Matrix questions
  - Fixed (control) and Random (treatment)
- Findings and Interpretations
  - Describes You as a Student
  - How Well Prepared
  - Important for Stanford to Provide
- Conclusions
- Discussion

Institutional Context

- Large research-intensive university, ~7,000 undergraduates
- Intense focus and reform in undergraduate education (SUES)
  - Owning knowledge
  - Adaptive learning
  - Redesigned freshman year


The Study of
Undergraduate Education
at Stanford University
January 2012

Survey Context

- Survey of New Students (SNS)
- Summer before start of freshman year
- ~1,650 invites
- ~80% response rate
- Very long survey
- Many matrix questions

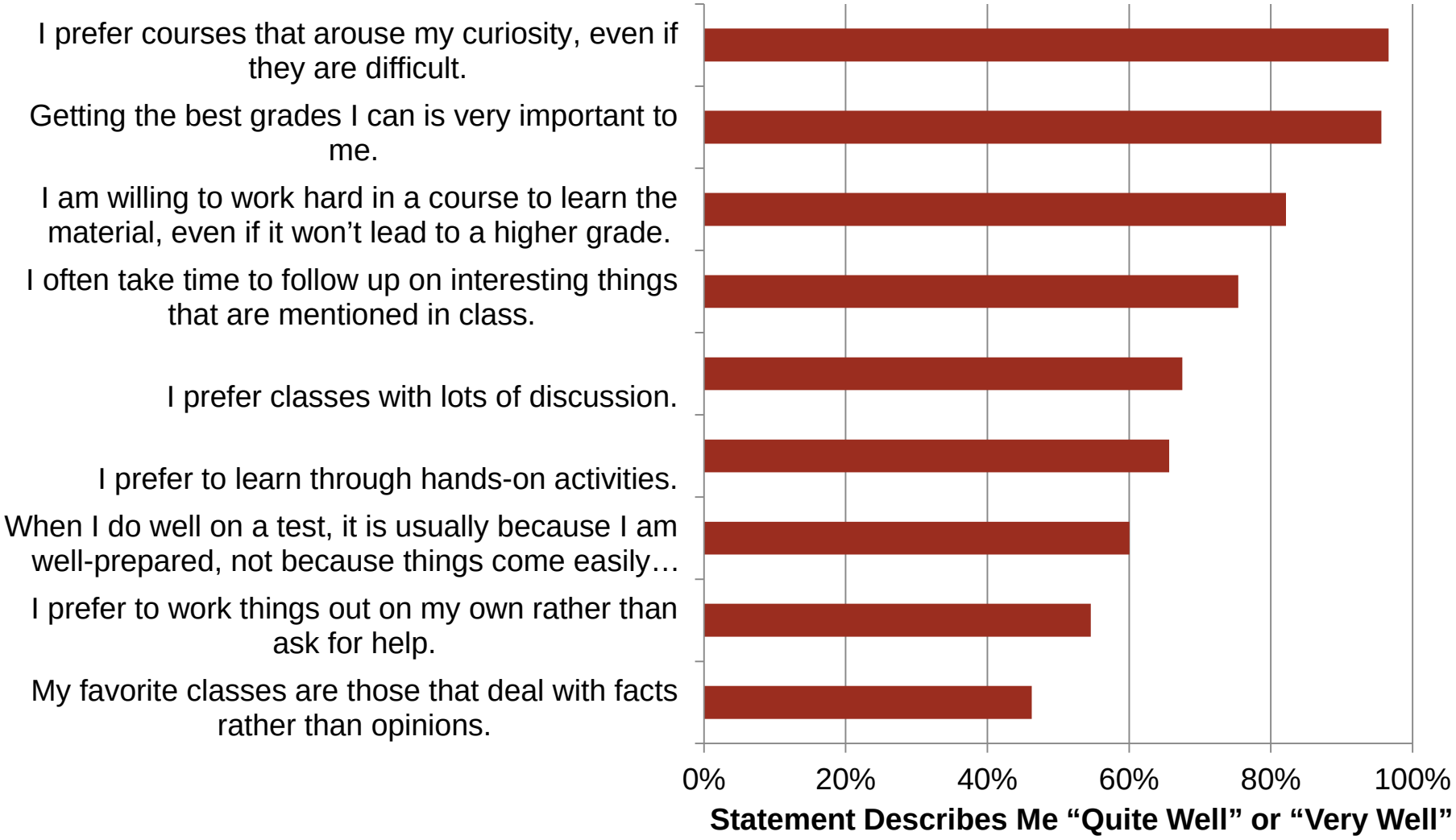
- Questions mostly inherited
- Desire to preserve trend and peer comparisons
- Faculty committees very invested
- Questions designed to feed policy decisions
- Use of individual items v. scale scores

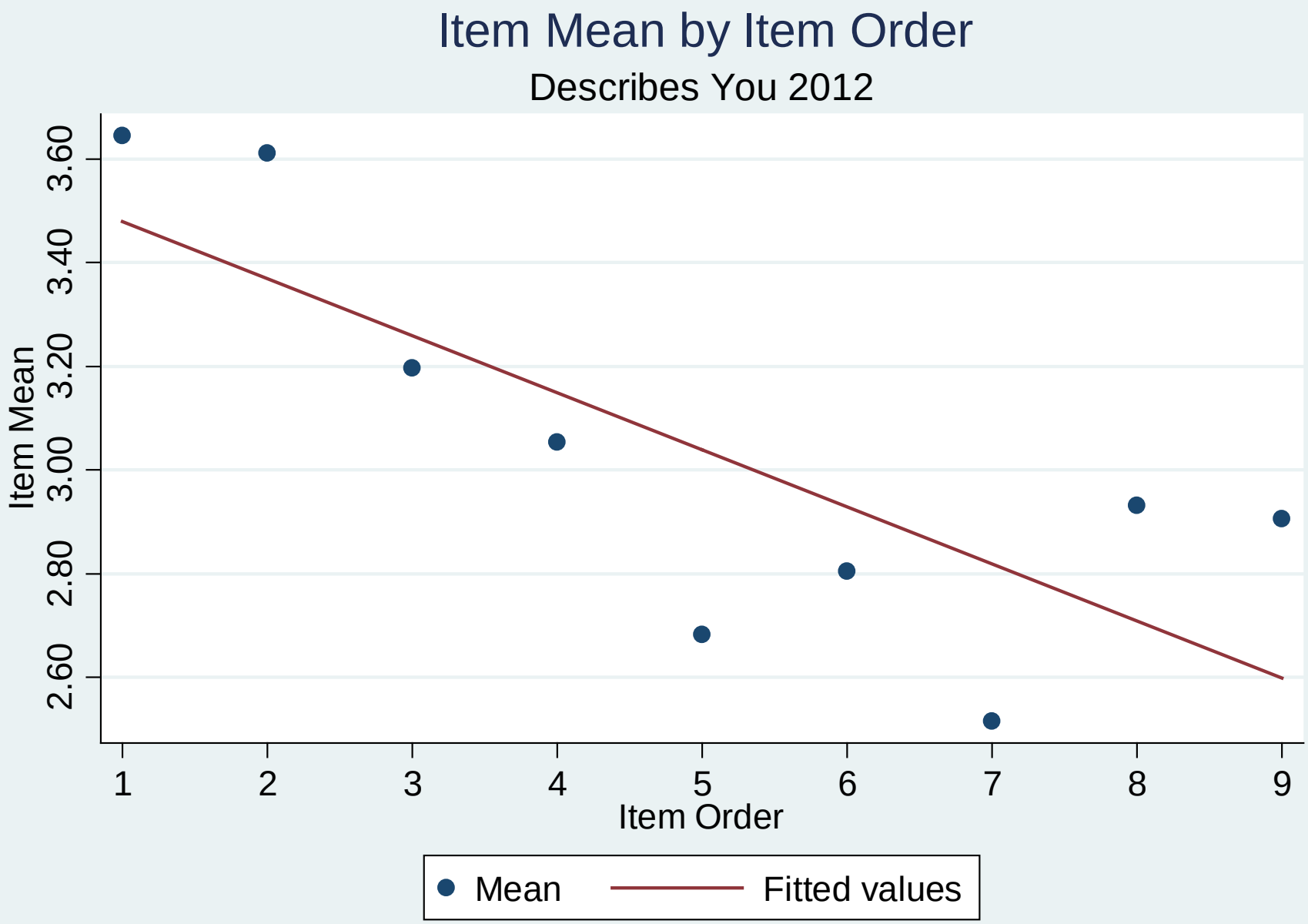
Motivation

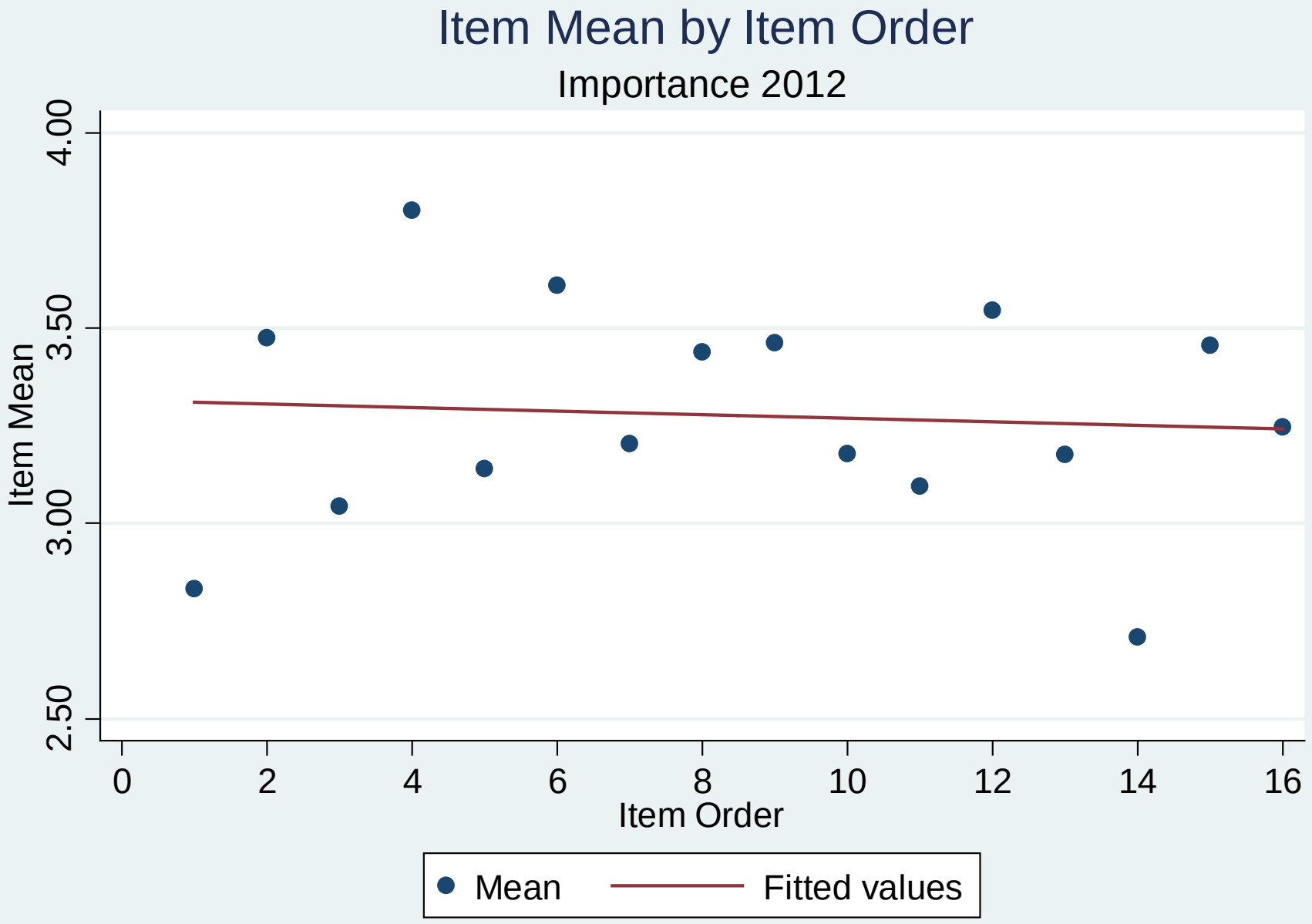
- Very strong apparent order effects in first matrix on survey
- Motivations and Learning Styles
- This matrix used heavily in reporting:
  - Faculty development
  - Student Affairs
  - Sub-group differences

- **Stakeholders love it, but is it real?**

2012 SNS
Describes You as a Student
Percent “Quite Well” or “Very Well”




Item Mean by Item Order
Describes You 2012
Item Mean
2.60
2.80
3.00
3.20
3.40
3.60
1
2
3
4
5
6
7
8
9
Item Order
Mean
Fitted values


Item Mean by Item Order
Importance 2012
Item Mean
4.00
3.50
3.00
2.50
0
2
4
6
8
10
12
14
16
Item Order
Mean
Fitted values

Prior Research

Why might we see order effects?

- Framing
  - Social Desirability
- Anchoring
- Survey fatigue

Study Details: Research Questions

For three policy-critical matrix questions from the Survey of New Students (SNS):

1. Are there relationships between item order and item results?
2. Are the substantive conclusions we draw about our students from these findings biased by item order?
3. If we believe there is bias, is randomizing items within a matrix a good solution?

Study Details: Activity

Please think back to the summer before your freshman year in college, and fill out the survey from that frame of mind.

Study Details: Matrix Questions

STANFORD UNIVERSITY

Class of 2017
Survey of New Students

2. How important is it that Stanford provide you with the following?

	Not important at all	Somewhat important	Very Important	Essential
A broad liberal arts education	○	○	○	○
Opportunities to develop skills valuable in the workforce	○	○	○	○
Contact with individuals whose backgrounds (e.g., race, socioeconomic status, nationality, sexual orientation) are different from your own	○	○	○	○
Opportunities to discover and pursue your intellectual passion	○	○	○	○

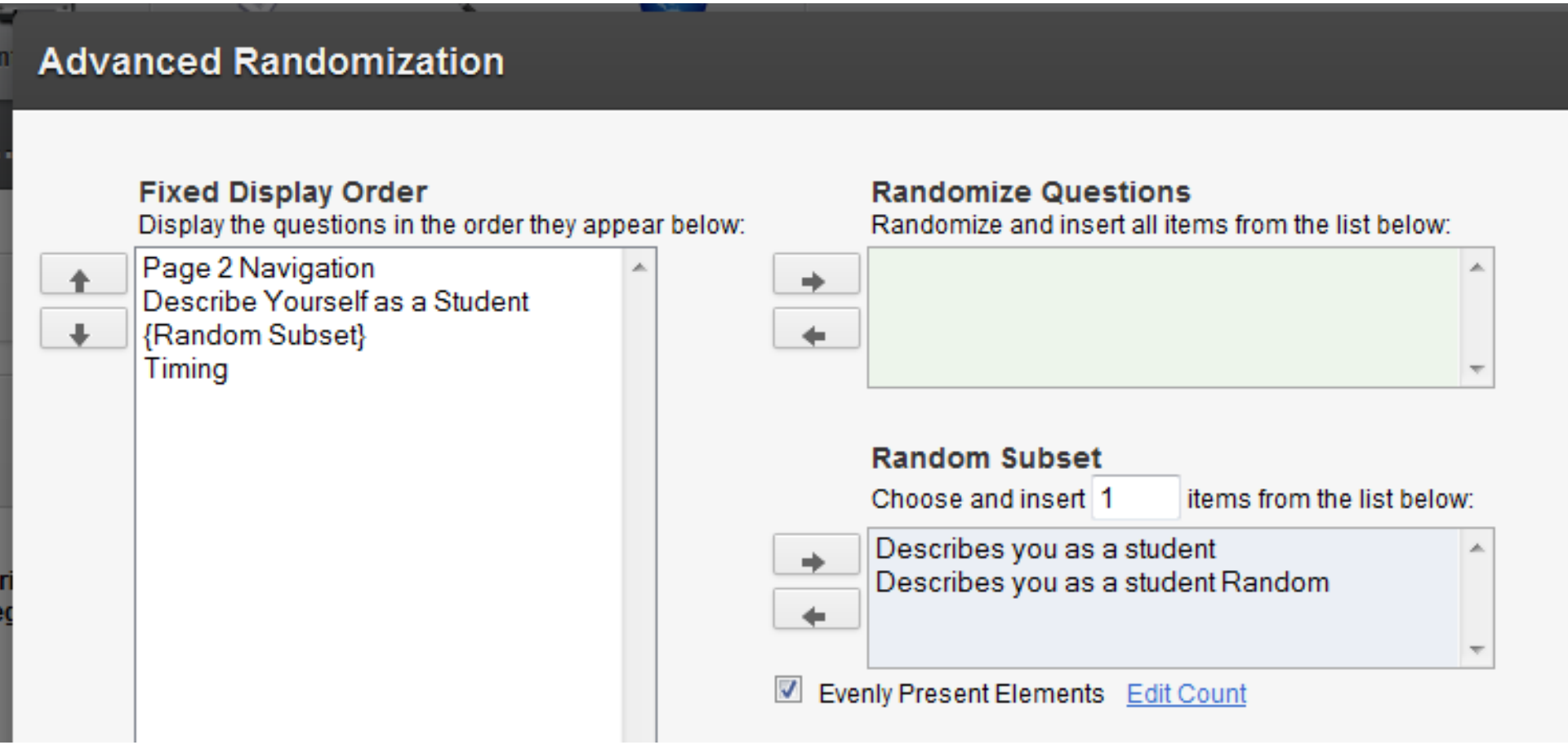
Study Details: Design

- Survey software random assignment
- Three separate experiments
  - ‘Control’ saw fixed order from prior years
  - ‘Treatment’ saw items randomized by survey software

Primary logic:

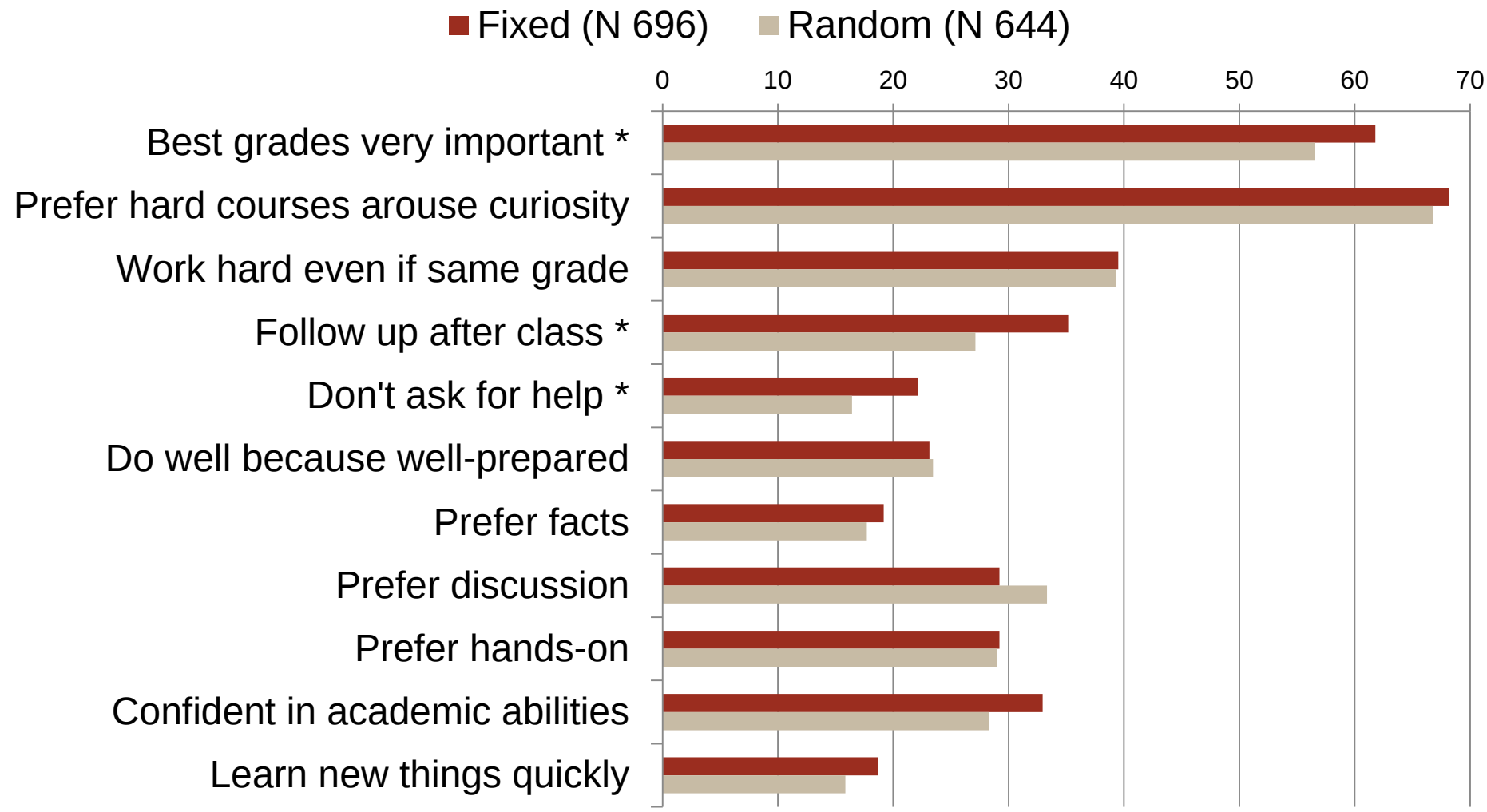
Difference in results between fixed and random?

Study Details: Qualtrics

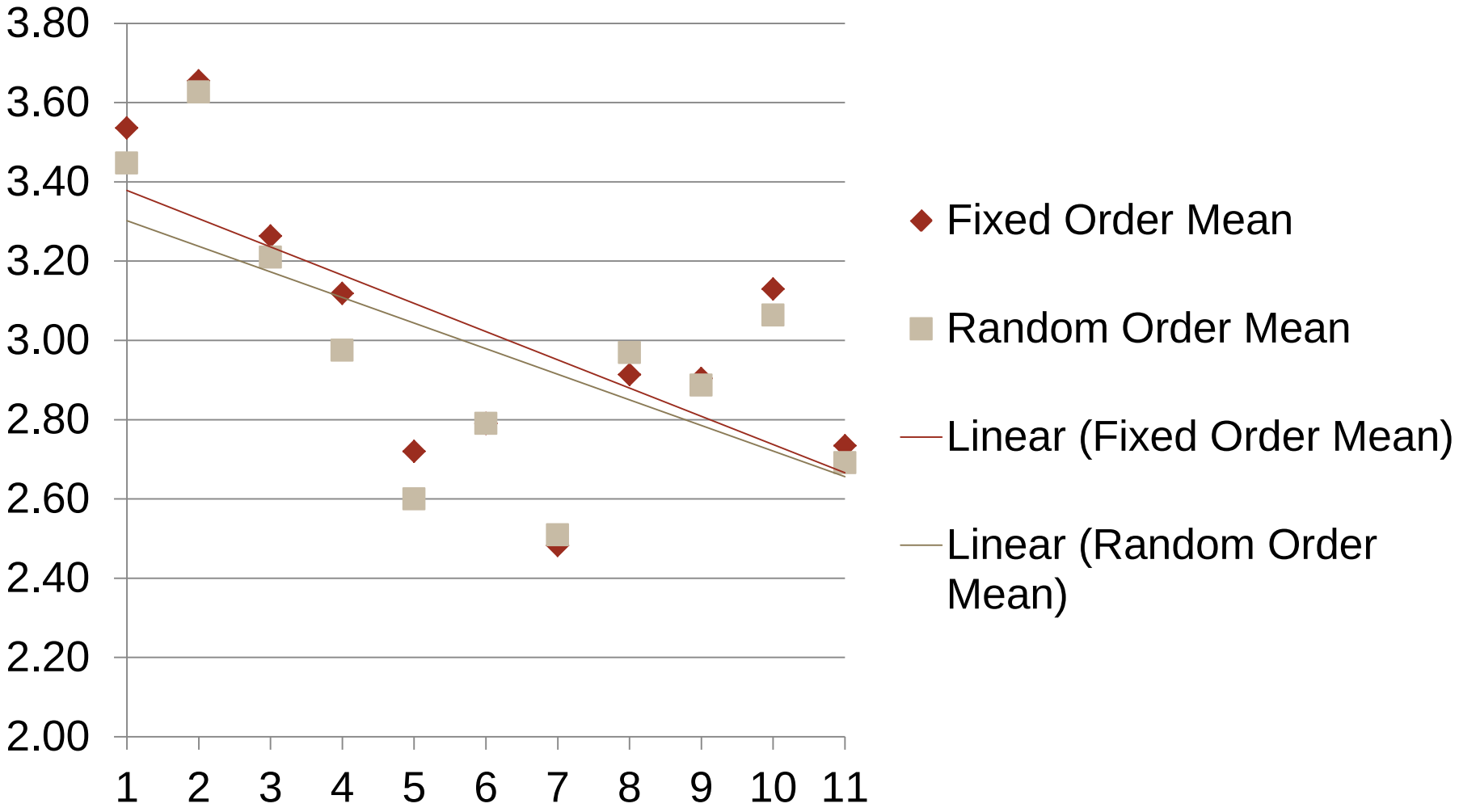


Findings and Interpretations

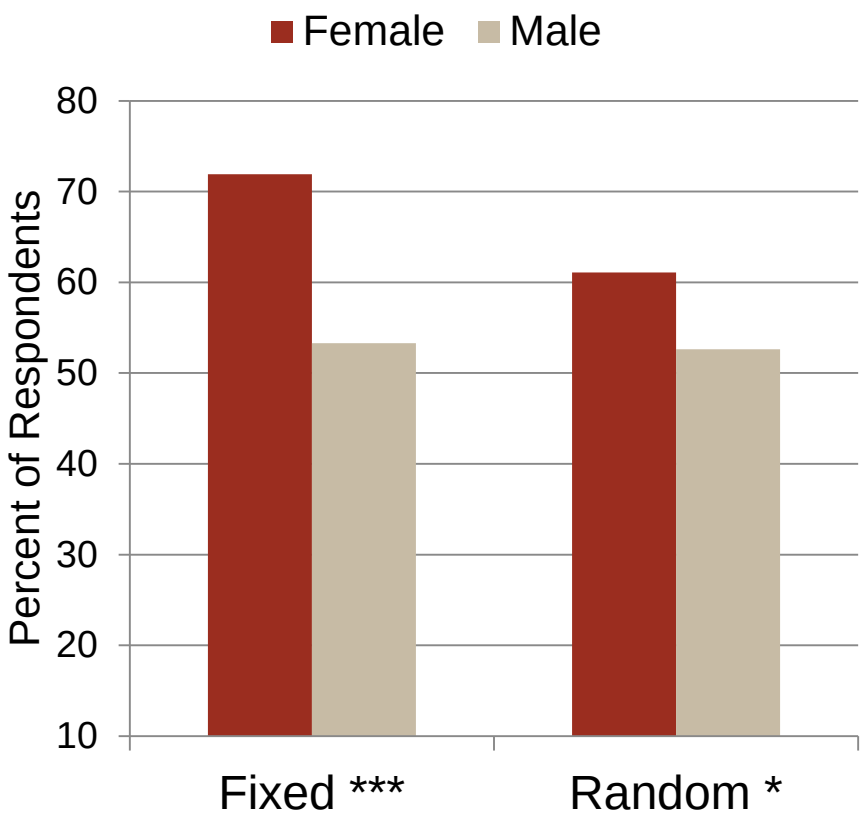
2013 Describes You as a Student
Fixed and Random Order
Percent "Very Well"



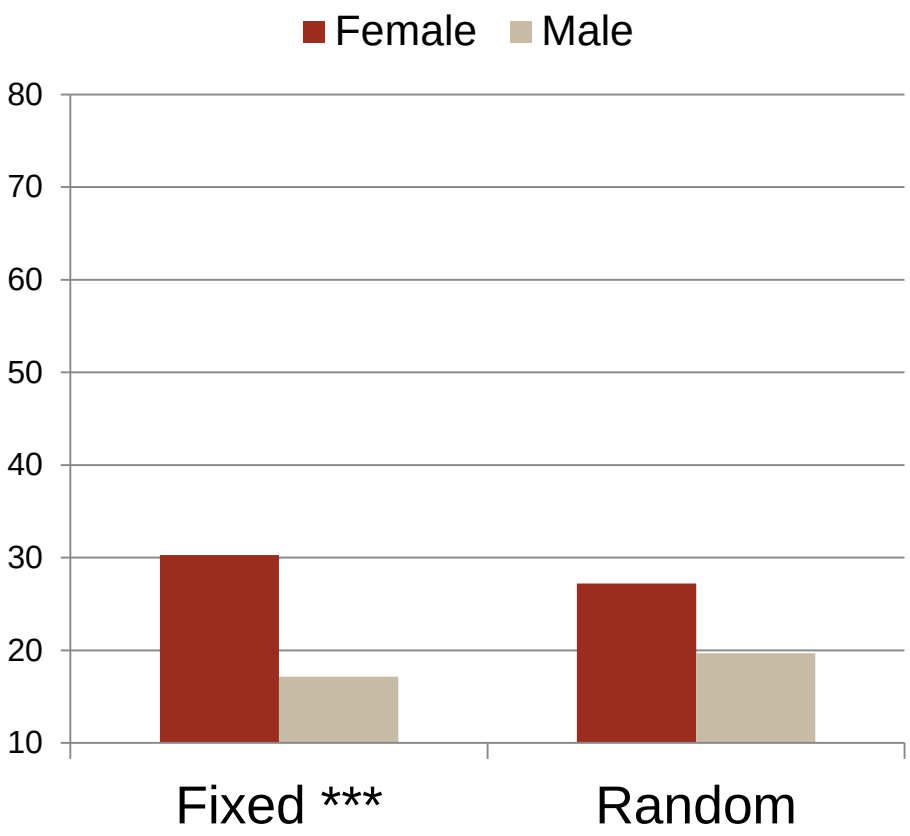
2013 Describes You as a Student
Fixed and Random Order
Item Mean by Item Order



2013 Best Grades Very Important
Fixed and Random Order
Gender Differences
Percent “Very Well”



2013 Do Well Because Well-Prepared
Fixed and Random Order
Gender Differences
Percent "Very Well"



Describes You as a Student Findings and Interpretations

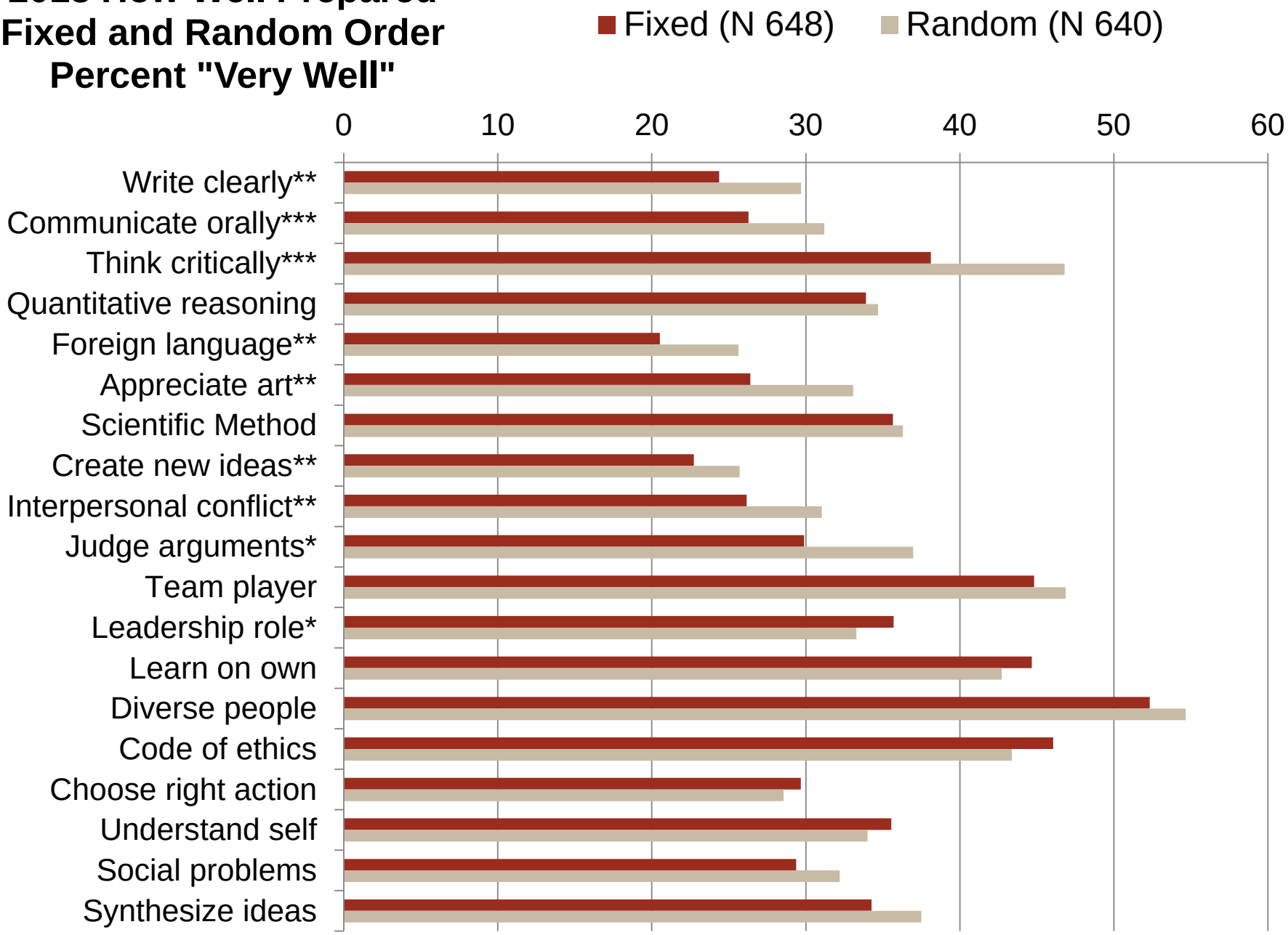
Findings

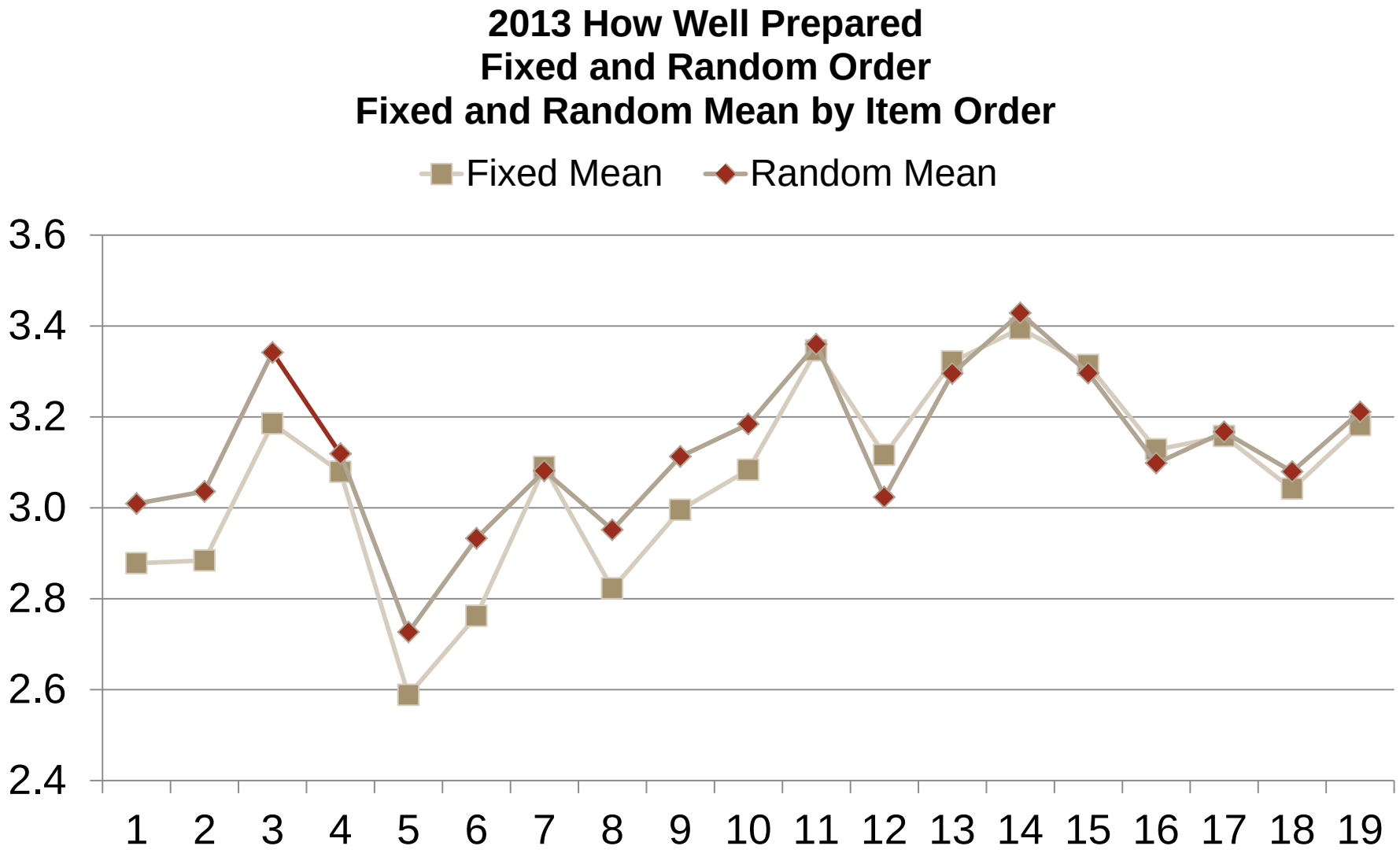
- No change in rank order of means
- Only a few statistically significant differences in item means
- Randomizing decreased percent “Very Well” 5-8% on some early items
- Less pronounced gender differences when randomized

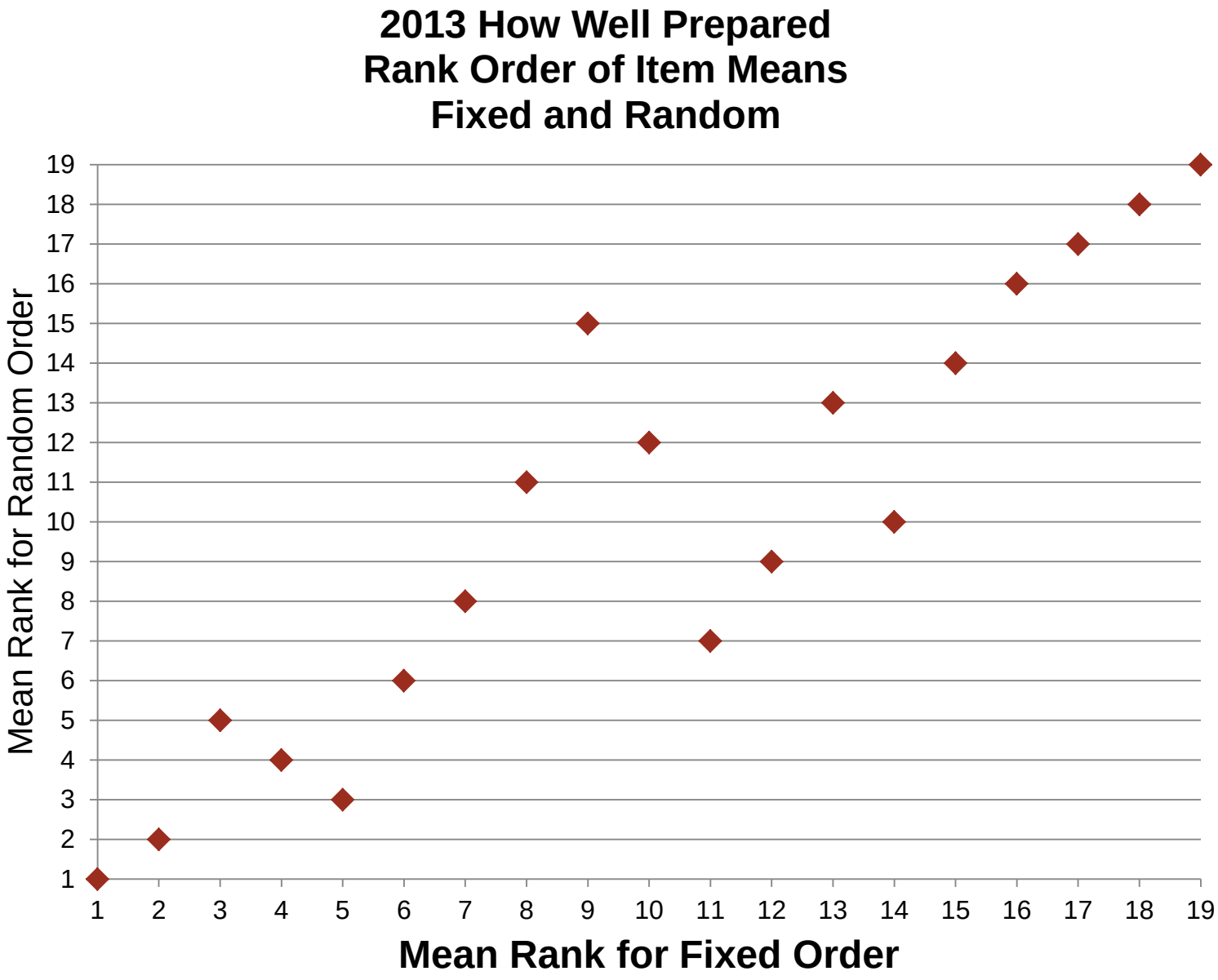
Interpretations

- Small item-order effects, but does not change the substantive interpretation of item comparisons
- Social desirability or enthusiasm in early items on the survey
- Interaction between social desirability or order effects and gender

2013 How Well Prepared Fixed and Random Order Percent "Very Well"




2013 How Well Prepared
Fixed and Random Order
Fixed and Random Mean by Item Order
Fixed Mean
Random Mean
3.6
3.4
3.2
3.0
2.8
2.6
2.4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19


2013 How Well Prepared
Rank Order of Item Means
Fixed and Random
Mean Rank for Random Order
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Mean Rank for Fixed Order

How Well Prepared
Findings and Interpretations

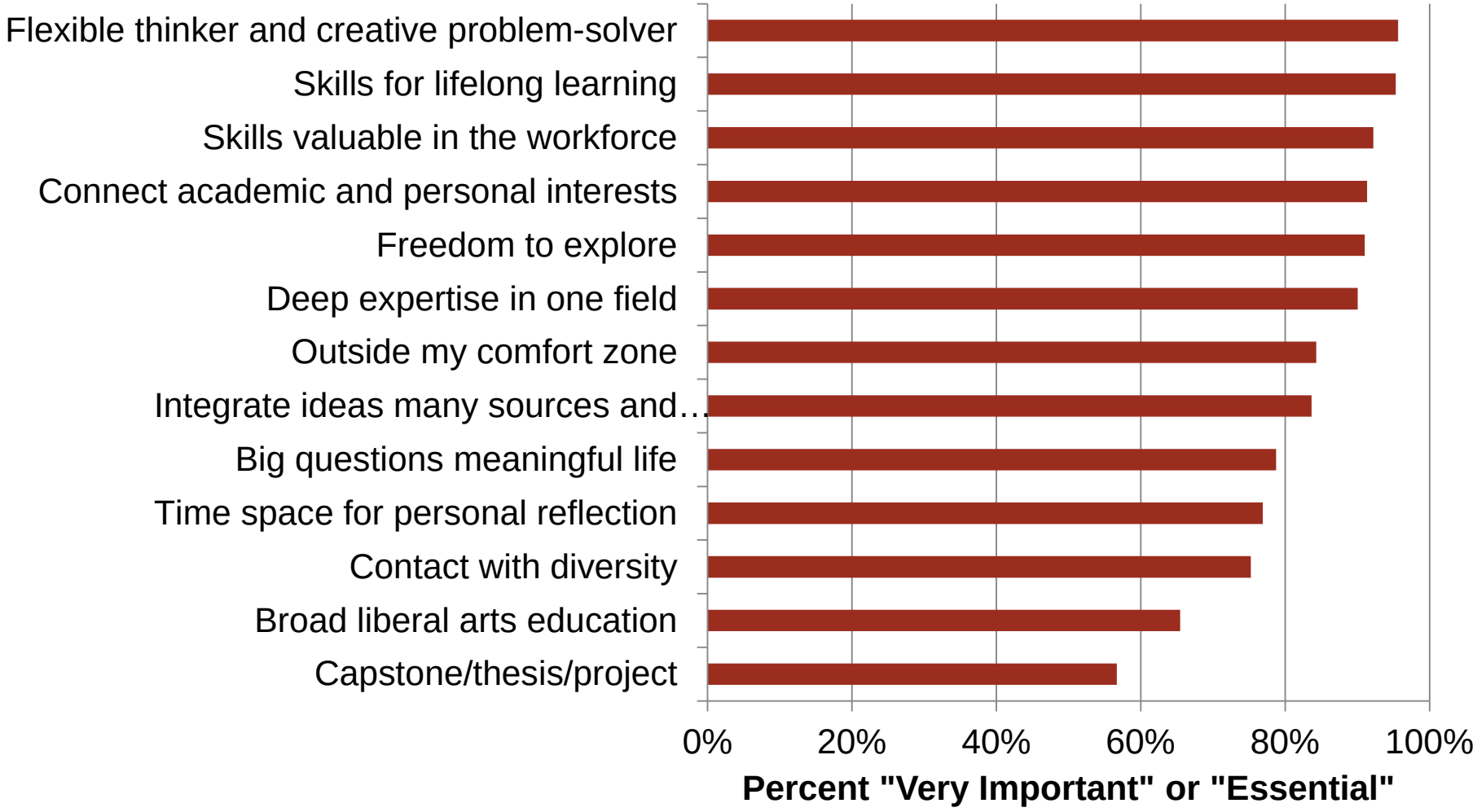
Findings

- 9 of 19 items have significantly different means
- 10 of 19 have different mean rank
- Differences much more pronounced in first half of matrix
- Quantitative and science showed no difference

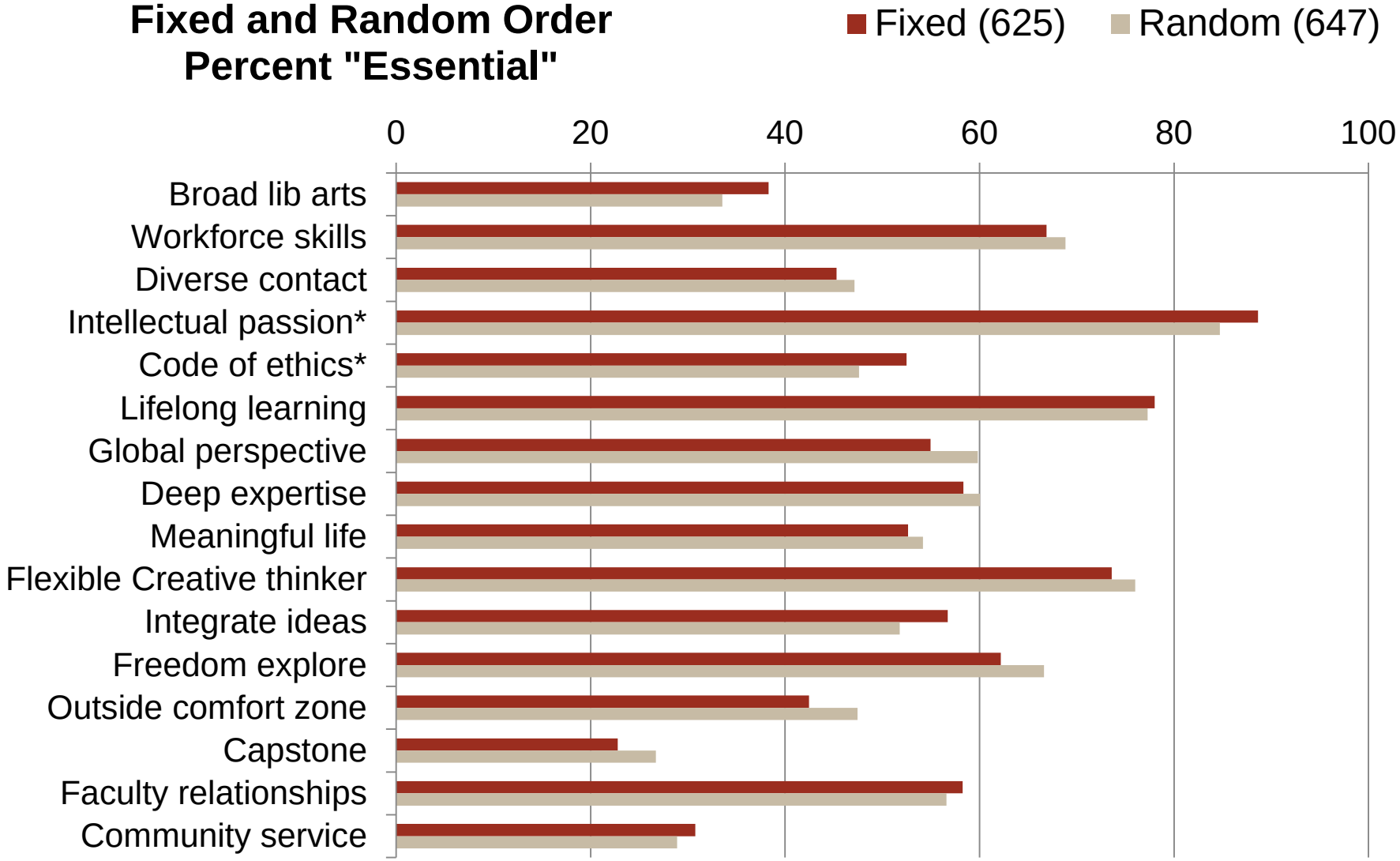
Interpretations

- Some kind of anchoring going on
- Threat to validity if items are clustered substantively, especially with respect to cross-item comparisons

2012 SNS
Importance of Stanford Providing...



2013 Importance of Stanford Providing Fixed and Random Order Percent "Essential"



Importance of Stanford Providing Findings and Interpretations

Findings

- Significant differences on only a couple of items
- All rank changes less than three

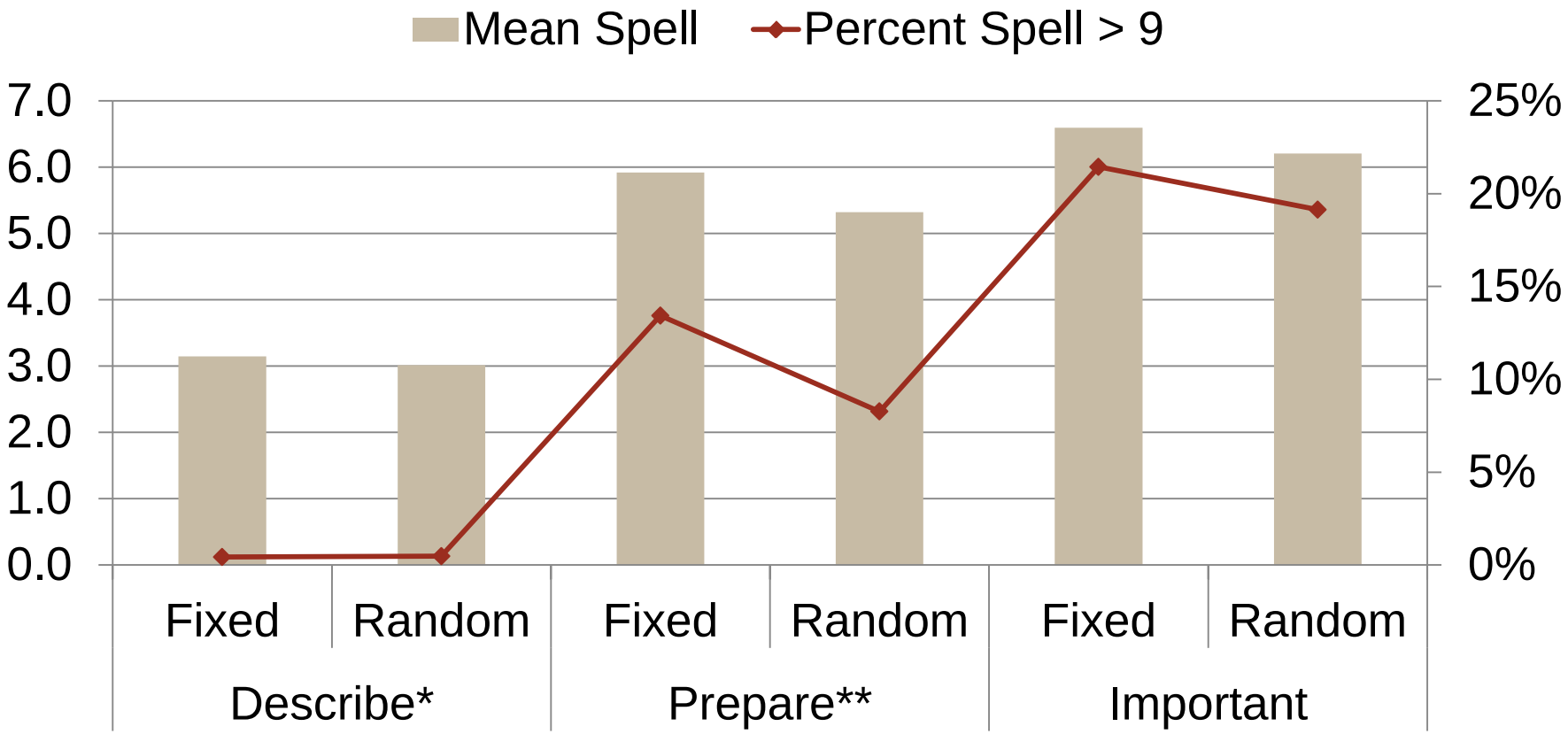
Interpretations

- Order effects don’t appear problematic

Concerns

- Satisficing
  - Non-differentiation (straight-line)
  - Drop in N

2013 SNS
Straight-Line Responding
Fixed and Random



Conclusions

How bad are they?

- Not too bad
- Strongest at
  - Beginning of survey
  - Beginning of matrix
  - Items grouped by content
  - Extreme item start of matrix (anchoring)

Impact on Substantive Conclusions

- Item comparisons
- Subgroup comparisons

Recommendation

- Randomize where possible
  - Trend, peers
- Break up substantive clusters
- If not randomize, put neutral questions first
- Monitor for straight-line reporting and satisficing