Respondent Burden: Lessons Learned from a Mixed-Methods Approach to Assessing Measures of Perceived Burden

Morgan Earp
Scott Fricker
John Dixon

BLS
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

Multi-Method Assessment Overview

- **Reliability (Internal Consistency)**
- **Distinguishability of Item Sub Scales**
- **Dimensionality**
  - Do the items measure the latent construct well?
  - Do they measure more than the latent construct?
- **Linearity**
  - Are the item sub scales linearly related to the latent variable?

BLS

Overview

- Potential sources of measurement error
- Item assessment tools
- Case study
  - Internal consistency
  - Dimensionality
  - Item Sub Scale Performance
    - Distinguishability
    - Linearity Summary
- Conclusions
- Next steps

BLS

Research Questions

- 1) Are responses to the items used in the measure internally consistent?

BLS

Research Questions

- 2) If not, why do they vary?
  - A) Are item relationships higher or lower than we would expect?

BLS

Research Questions

- 3) How are the subscales of my measure performing?
  - A) Do the different response options distinguish between varying levels of latent burden?
  - B) Are the ordinal response options linearly related to each other and latent burden as whole?

BLS

Models & Statistics

- **Confirmatory Factor Analysis**
  - Error Variance (internal consistency)
- **Item Response Theory**
  - Chi Square (internal consistency)
  - LD Chi Square (dimensionality)
  - Trace-lines (distinguishability of sub-scales)
  - Mean Square Error (linearity of sub-scales)

BLS

Assessed 9 Burden Items

- Burden
- Length
- Interest
- Difficulty
- Sensitivity
- Appropriate # of Waves
- Another Survey
- Extend Survey
- Effort

BLS

Background

- Current Expenditure Survey (CEQ)
  - respondents were asked about burden in a post survey (end of 5th interview)
- Burden items were developed using Bradburn's (1978) definition:
  - effort
  - survey length
  - frequency of survey requests

BLS

Research Question 1

- 1) Are responses to the items used in the measure internally consistent?

BLS

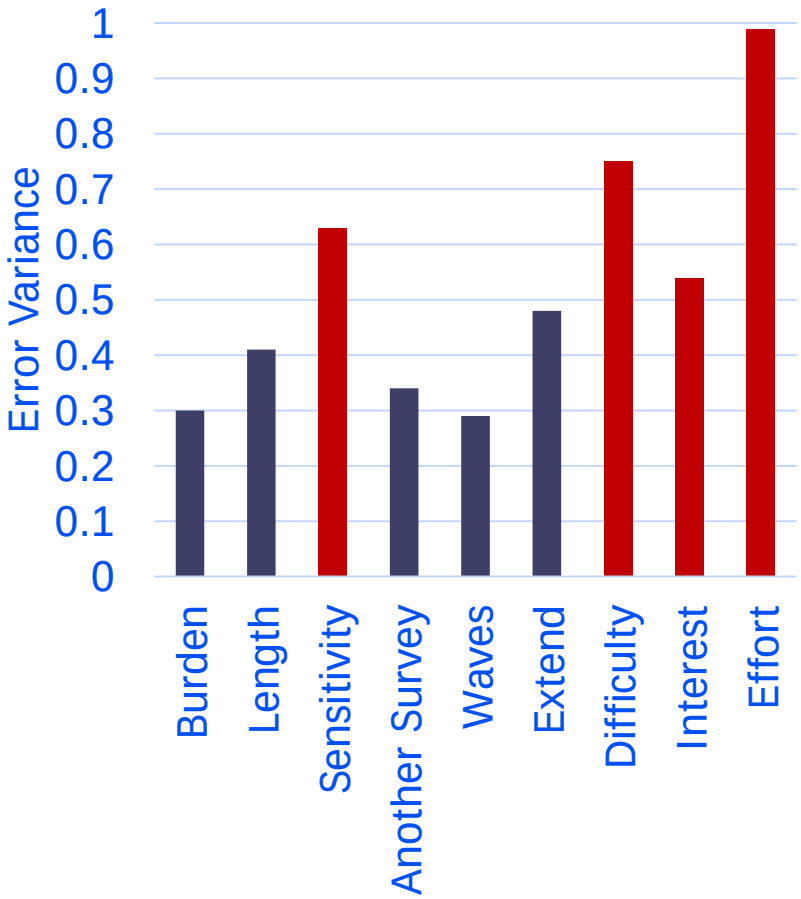
Internal Consistency

- Confirmatory Factor Analysis, Error Variance: $1 - r^2$
  - Measures the *% of variation* in the latent construct not explained by the item
- Item Response Theory – Chi Square Statistic
  - Is there a significant proportion of variation in the latent construct not explained by this item?

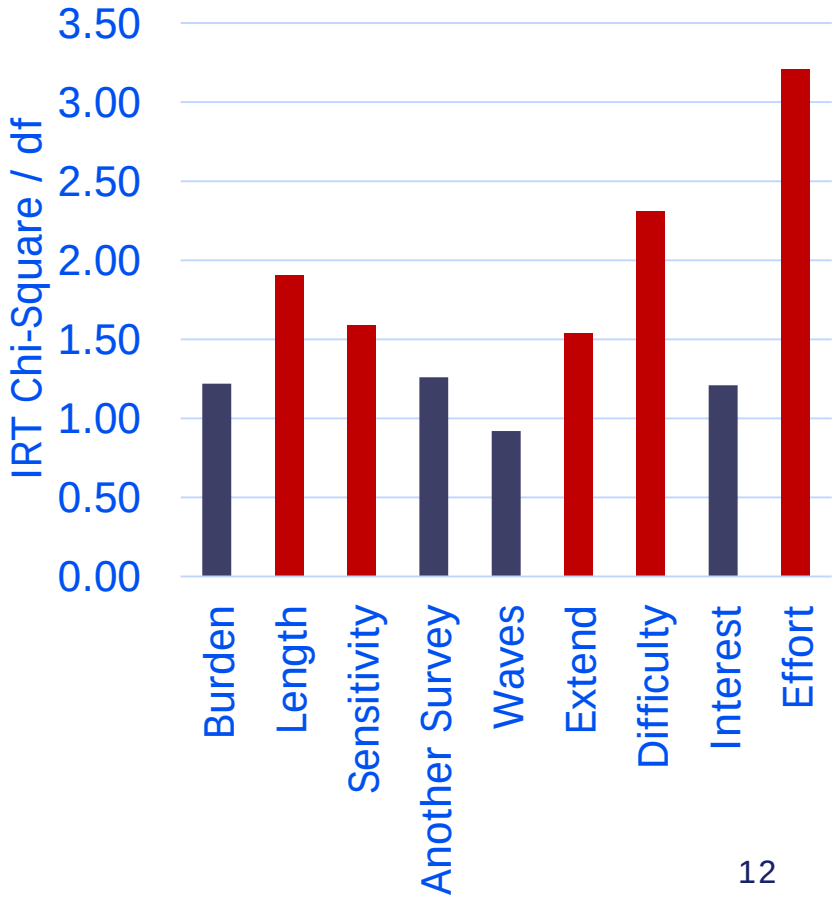
BLS

Internal Consistency

CFA – Error Variance



IRT – Chi Square



BLS

Item Response Theory - Chi Square

- X^2 also considers
  - how difficult an item is to agree w/
    - Some items maybe carry more weight…
  - Item response patterns vs total score
  - local independence of the item
    - Is it as correlated as we would expect w/ the other items?
    - Is it more correlated w/ the other items than we would expect after controlling for the latent variable burden?

BLS

Research Question 2

- 2) If item responses are not internally consistent, why might they vary?
  - A) Are item relationships higher or lower than we would expect?
  - B) Are there interaction effects aka differential item functioning?

BLS

Local Independence

- Dimensionality
  - LD $\chi^2 \leq -10$
    - not as correlated as we would expect w/ another item
  - LD $\chi^2 \geq 10$
    - more correlated than we would expect w/ another item after controlling for the latent burden
    - measures more than just latent burden

BLS

IRT LD χ^2 Statistics by Burden Items

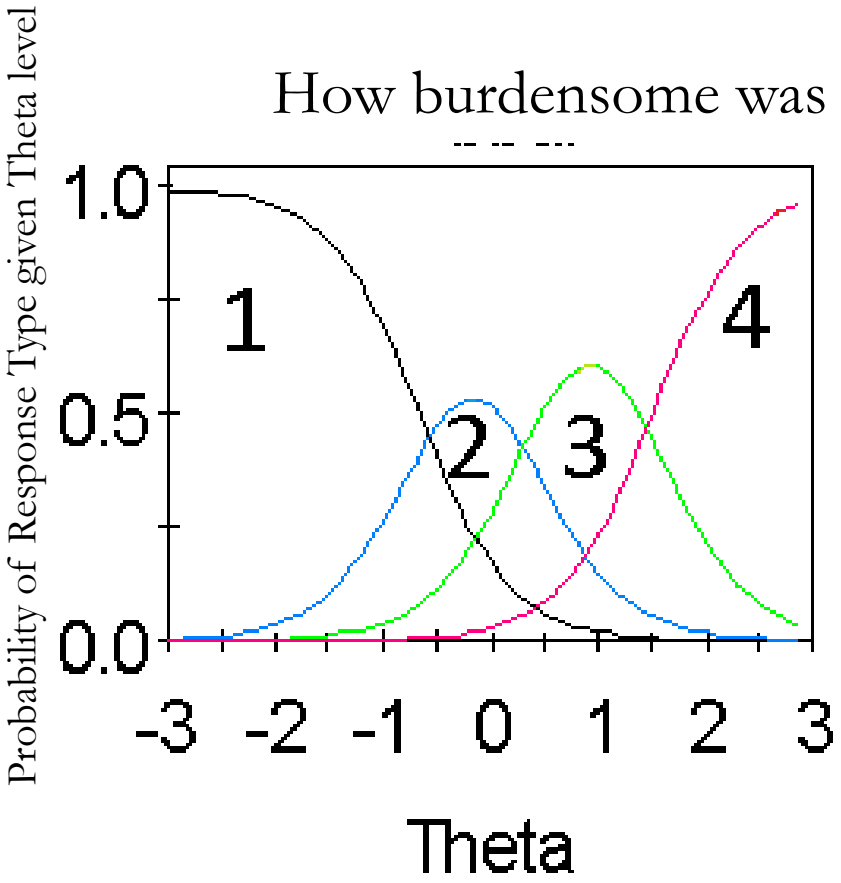
Item	Burden	Another	Extend	Length	Interest	Sensitivity	Wave	Difficulty
Burden								
Another	-7.7							
Extend	-10.1	-15.0						
Length	-8.6	7.6	5.0					
Interest	-17.5	-5.4	-7.0	-0.9				
Sensitivity	14.9	7.8	-9.5	-4.2	-17.2			
Wave	-3.4	-0.7	-0.0	0.2	0.1	-8.6		
Difficulty	26.6	12.3	-11.9	1.4	-13.8	18.3	-8.8	
Effort	20.0	7.6	-12.9	1.9	-28.7	8.1	-0.1	13.1

BLS

Research Question 3

- 3) How are the subscales of my measure performing?
  - A) Do the different response options distinguish between varying levels of latent burden?

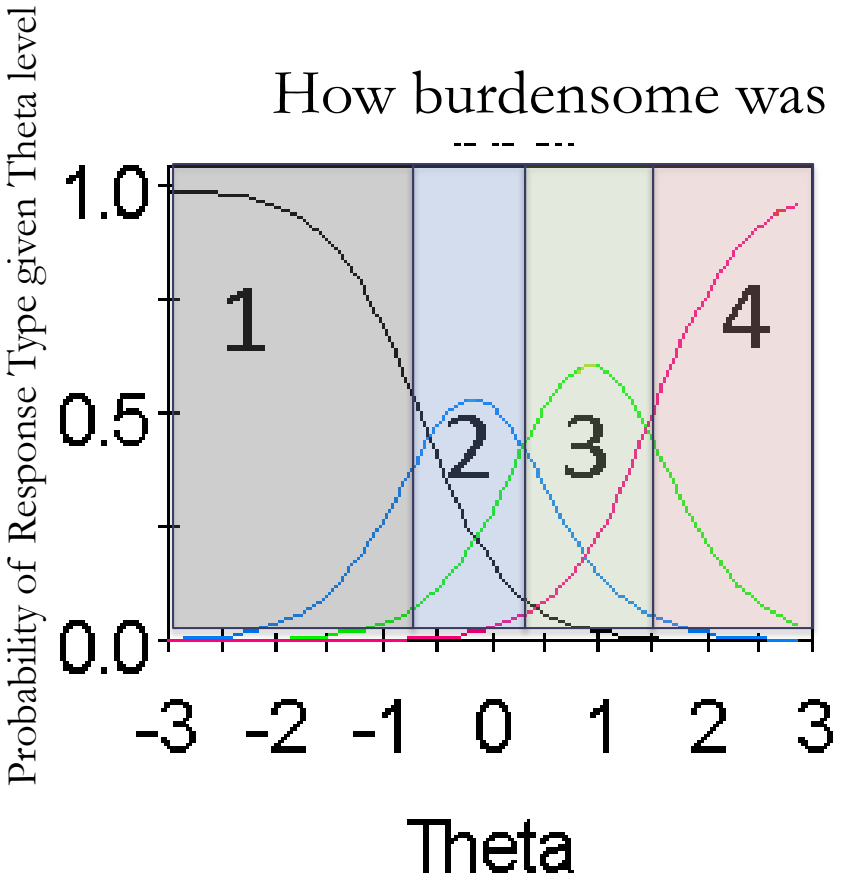
Tracelines
Good Item Fit Example



1 Not at all burdensome
2 A little burdensome
3 Somewhat burdensome
4 Very burdensome

Theta = Standardized Scale of the Latent Variable Burden

Tracelines
Good Item Fit Example

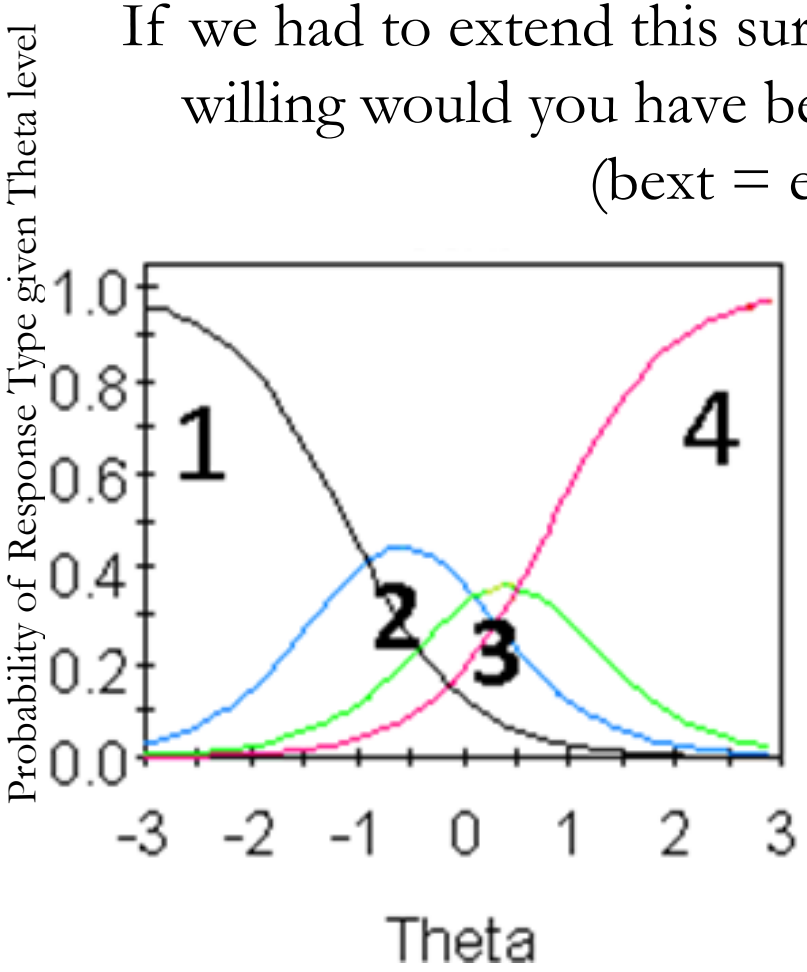


1 Not at all burdensome
2 A little burdensome
3 Somewhat burdensome
4 Very burdensome

Theta = Standardized Scale of the Latent Variable Burden

Tracelines
Poor Item Fit Example

If we had to extend this survey for another 15 minutes, how willing would you have been to continue this interview? (bext = extend survey)



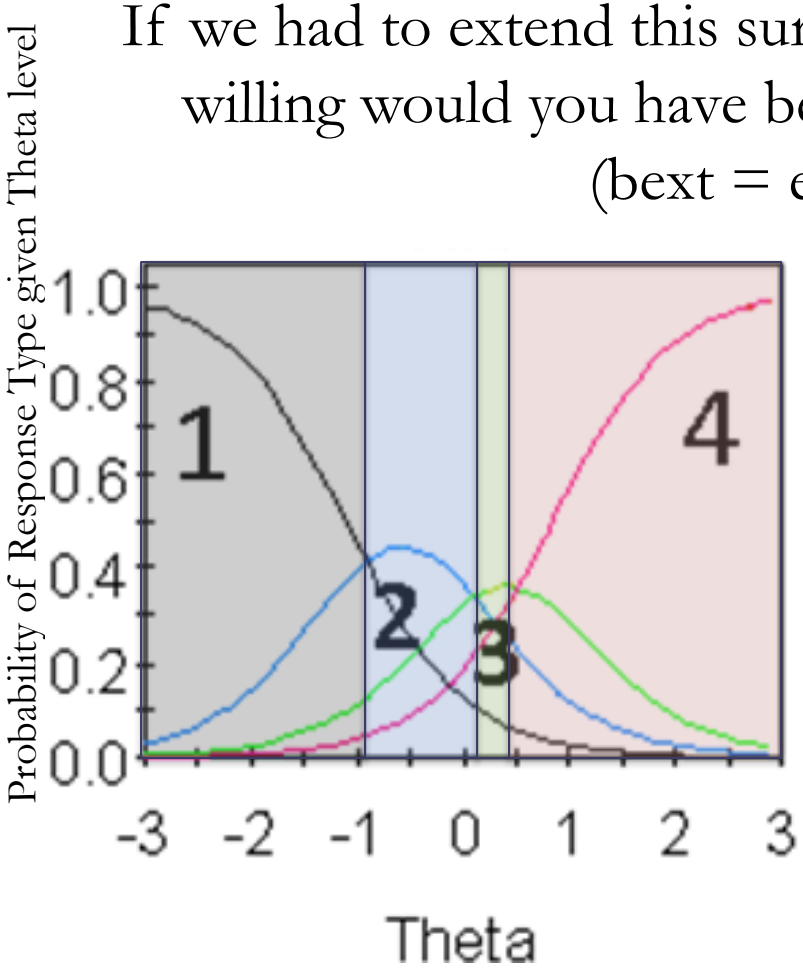
1 Very willing
2 Somewhat willing
3 Somewhat unwilling
is not associated w/ unique levels of burden
4 Very unwilling

Theta = Standardized Scale of the Latent Variable Burden

BLS

Tracelines
Poor Item Fit Example

If we had to extend this survey for another 15 minutes, how willing would you have been to continue this interview? (bext = extend survey)



1 Very willing
2 Somewhat willing
3 Somewhat unwilling
is not associated w/ unique levels of burden
4 Very unwilling

Theta = Standardized Scale of the Latent Variable Burden

Research Question 3

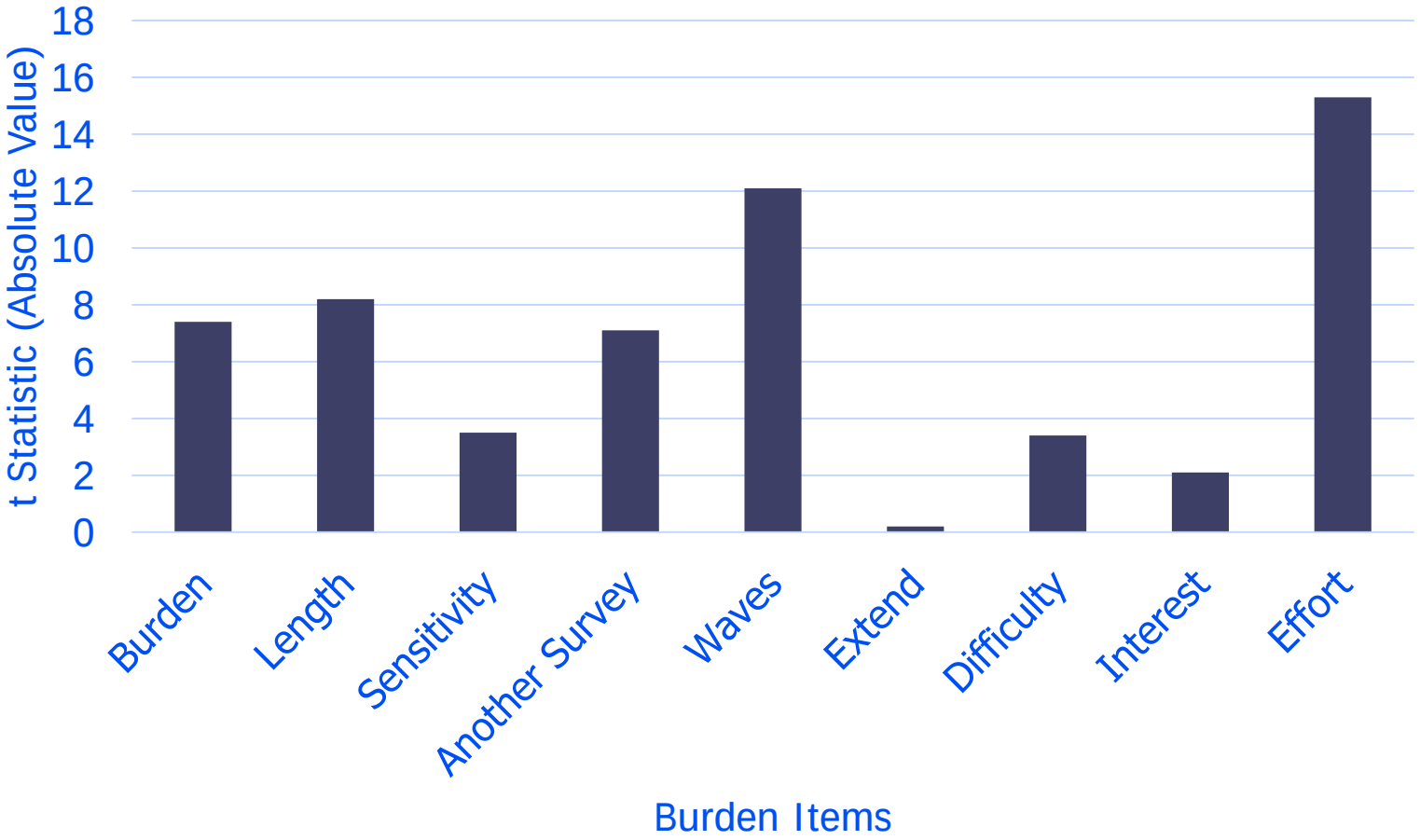
- 3) How are the subscales of my measure performing?
  - A) Do the different response options distinguish between varying levels of latent burden?
  - B) Are the ordinal response options linearly related to each other and latent burden as whole?

IRT - Linearity

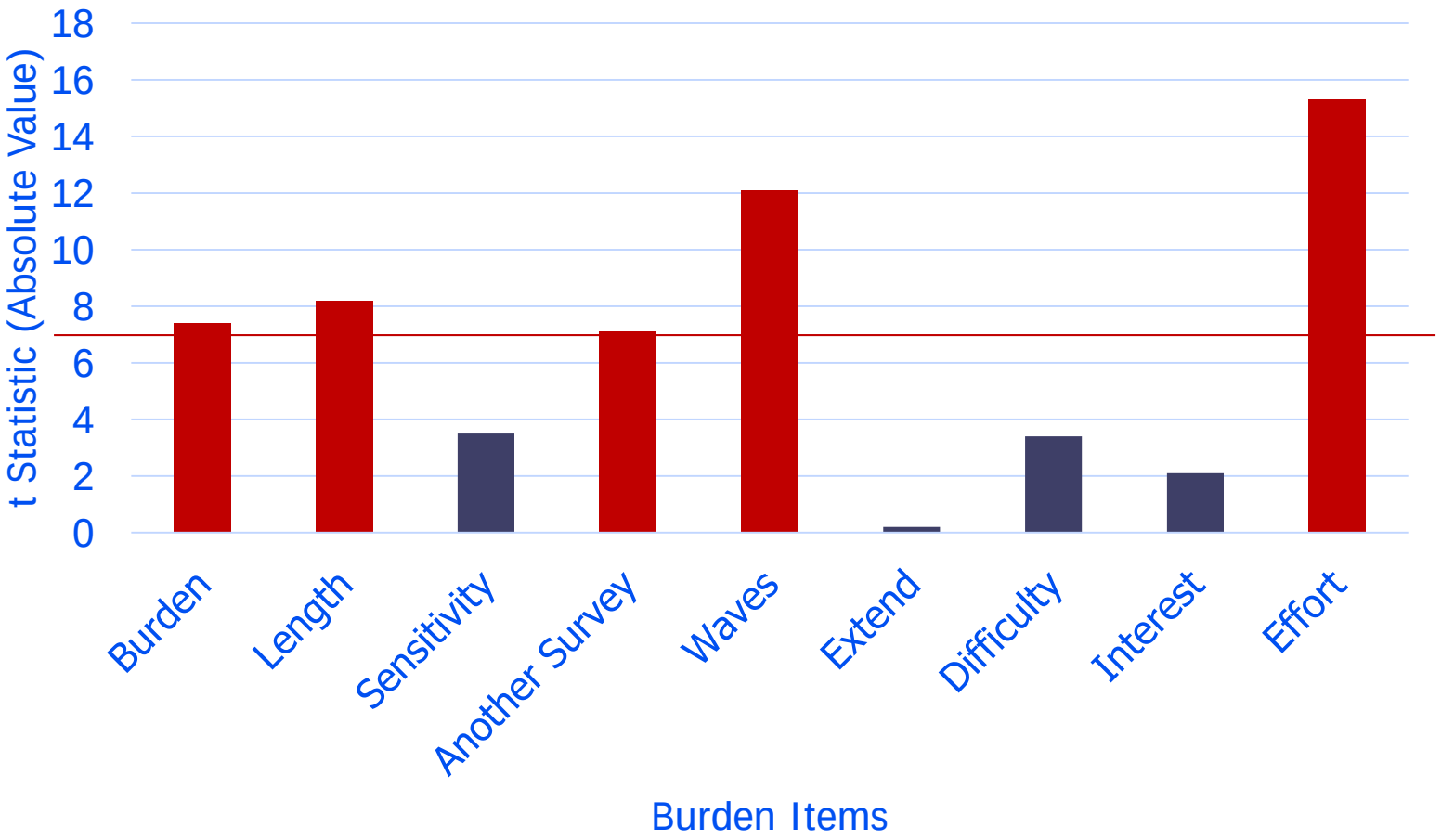
- Linearity (Mean Square Error: *t*)
  - Like CFA and the IRT Chi-Square statistic, the IRT *t* statistic also looks at the consistency of item responses, but is more sensitive to violations of linearity
  - If items responses violate the assumptions of linearity (they do not have a linear relationship w/ the latent variable), they will have significant misfit

BLS

t Statistics by Burden Items

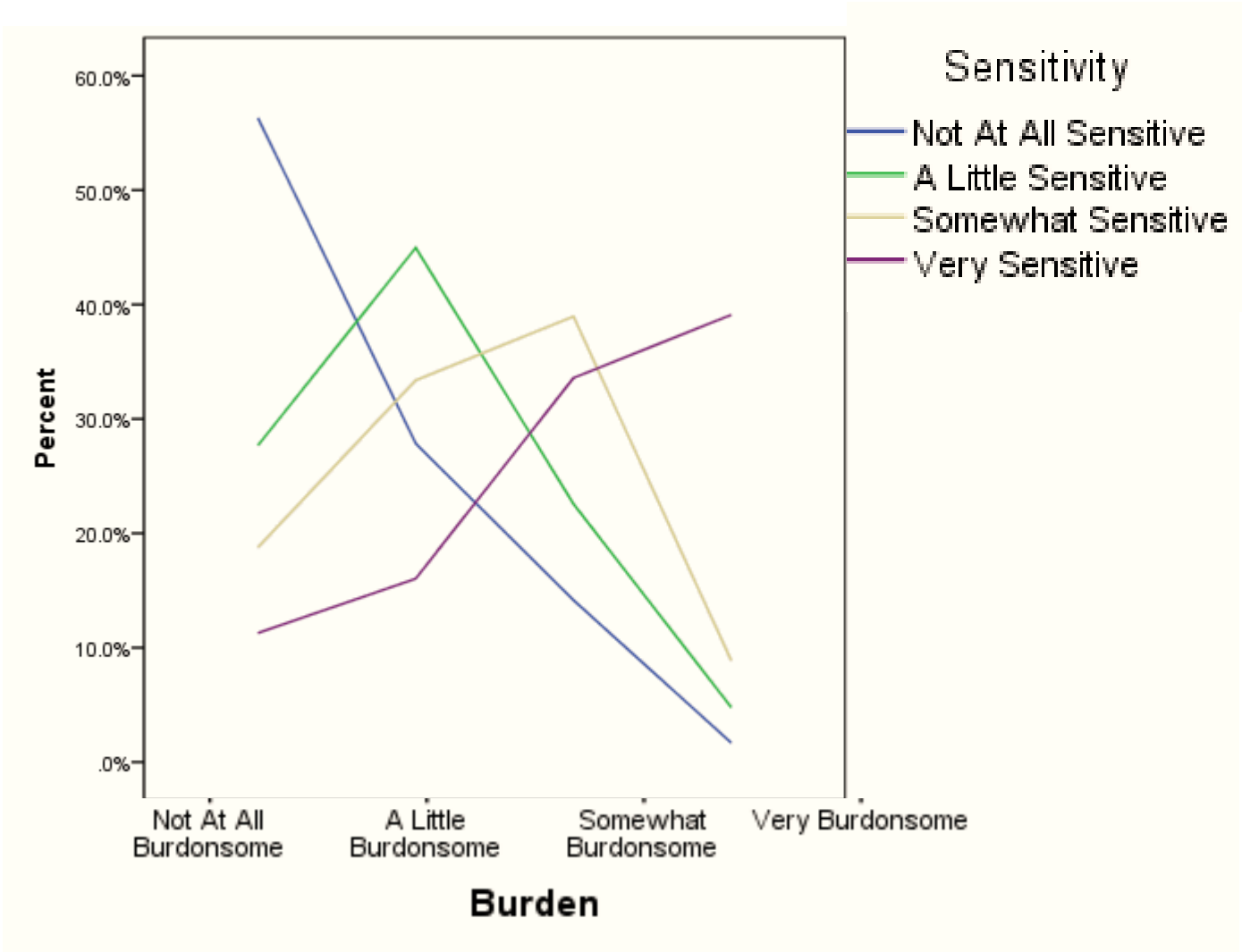


t Statistics by Burden Items

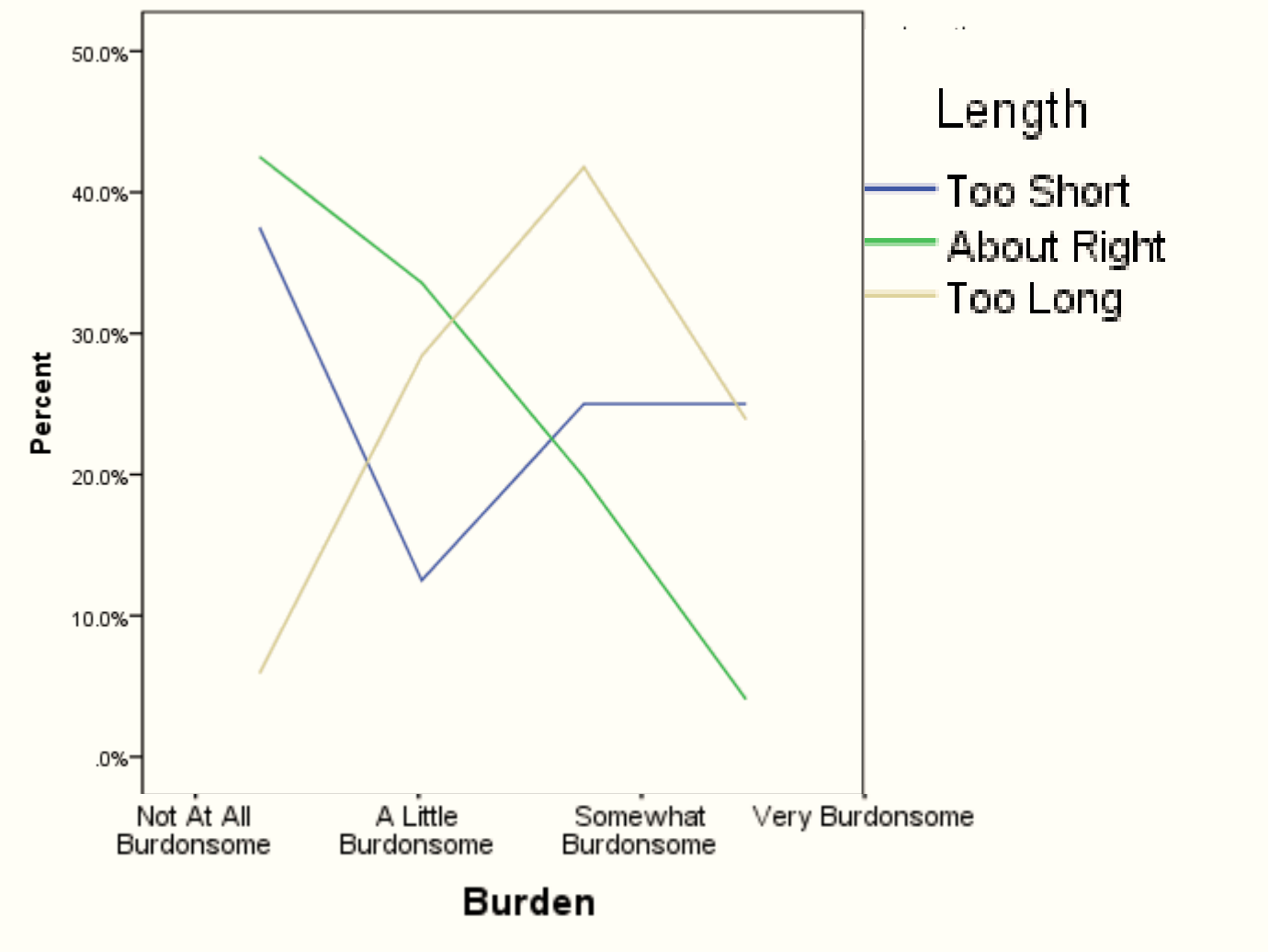


Significant Misfit

Item Response Curves by Reported Burden Good Fitting Item



Item Response Curves by Reported Burden Poor Fitting Item



Conclusions

- While psychometric assessment tools provide insight about how our items perform - There is no single cohesive tool – it takes multiple methods
- All methods provide different details (even w/in IRT) – item may perform well in one aspect, but not another
- The various methods can be used to better understand and potentially improve our items

BLS

Our Next Steps

- Exploring how to best improve item sub-scales to improve utility and linearity across all response options

- Examining our definition of burden and whether we should include more items to measure a multidimensional construct or reduce items to assess a single dimensional construct

BLS

Contact Information

Contact Information

Morgan Earp

Earp.Morgan@bls.gov

BLS

BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR