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

Morgan Earp
Scott Fricker
John Dixon
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?
Overview

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

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

2) If not, why do they vary?

A) Are item relationships higher or lower than we would expect?
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?
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)
Assessed 9 Burden Items

- Burden
- Length
- Interest
- Difficulty
- Sensitivity

- Appropriate # of Waves
- Another Survey
- Extend Survey
- Effort
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
Research Question 1

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

CFA – Error Variance

IRT – Chi Square

Error Variance

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Burden  Length  Sensitivity  Another Survey  Waves  Extend  Difficulty  Interest  Effort

IRT Chi-Square / df

0.00
0.50
1.00
1.50
2.00
2.50
3.00
3.50

Burden  Length  Sensitivity  Another Survey  Waves  Extend  Difficulty  Interest  Effort
Item Response Theory - Chi Square

$X^2$ also considers:

- how difficult an item is to agree with
  - 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 with the other items?
  - Is it more correlated with the other items than we would expect after controlling for the latent variable burden?
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?
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
# IRT LD $\chi^2$ Statistics by Burden Items

<table>
<thead>
<tr>
<th>Item</th>
<th>Burden</th>
<th>Another</th>
<th>Extend</th>
<th>Length</th>
<th>Interest</th>
<th>Sensitivity</th>
<th>Wave</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burden</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Another</td>
<td>-7.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extend</td>
<td>-10.1</td>
<td>-15.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>-8.6</td>
<td>7.6</td>
<td>5.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest</td>
<td>-17.5</td>
<td>-5.4</td>
<td>-7.0</td>
<td>-0.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>14.9</td>
<td>7.8</td>
<td>-9.5</td>
<td>-4.2</td>
<td>-17.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave</td>
<td>-3.4</td>
<td>-0.7</td>
<td>-0.0</td>
<td>0.2</td>
<td>0.1</td>
<td>-8.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulty</td>
<td>26.6</td>
<td>12.3</td>
<td>-11.9</td>
<td>1.4</td>
<td>-13.8</td>
<td>18.3</td>
<td>-8.8</td>
<td></td>
</tr>
<tr>
<td>Effort</td>
<td>20.0</td>
<td>7.6</td>
<td>-12.9</td>
<td>1.9</td>
<td>-28.7</td>
<td>8.1</td>
<td>-0.1</td>
<td>13.1</td>
</tr>
</tbody>
</table>
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?
How burdensome was this survey to you?

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

How burdensome was this survey to you?

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
4 Very unwilling

Theta = Standardized Scale of the Latent Variable Burden
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
4 Very unwilling

Probability of Response Type given Theta level

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.
t Statistics by Burden Items

Burden Items

- Burden
- Length
- Sensitivity
- Another Survey
- Waves
- Extend
- Difficulty
- Interest
- Effort

Values on the y-axis represent the absolute value of the t-statistic.
$t$ Statistics by Burden Items

![Bar chart showing t Statistic (Absolute Value) for various Burden Items, with 'Significant Misfit' indicated.](image-url)
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
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
Contact Information

Morgan Earp
Earp.Morgan@bls.gov