Understanding and Detecting Careless Responding in Survey Research

Acknowledgements: MK Ward, Paul Curran, Jason Huang, Nathan Bowling
Outline

- CR defined
- Importance of CR
- Causes of CR
- Ways to Address CR
  - CR Indicators
  - CR Prevention
Careless Responding (CR) Defined

Sometimes survey respondents (intentionally or unintentionally) respond to the survey in a manner that does not accurately reflect their true beliefs or feelings

– Content responsive
  • E.g., faking, social desirability, impression management

– Content non-responsive
  • Careless responding (insufficient effort responding, random responding, etc.)

• Careless responding occurs when responses to survey items are not based on the item content
Patterns of Response

• The problem with careless response….it can take many different forms
  – Random
  – Consistent response
  – Midpoint or extreme
  – Pervasive vs intermittent
• There is no single method to detect all possible types of careless response
<table>
<thead>
<tr>
<th>Random ordered items on survey</th>
<th>Uniform Random</th>
<th>Mid-point Random</th>
<th>Consistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>I love to eat</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I let others worry about higher values; my main concern is with the</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>bottom line</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The reason I vote is because my vote can make a difference</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>I like to look at my body</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I use others for my own ends</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>I often make last-minute plans</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I often admire a really clever scam</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I excel in what I do</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I get angry easily</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>I radiate joy</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Carelessness is often identified by a lack of consistency

<table>
<thead>
<tr>
<th>Friendliness</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make friends easily.</td>
<td>5</td>
</tr>
<tr>
<td>Warm up quickly to others.</td>
<td>5</td>
</tr>
<tr>
<td>Feel comfortable around people.</td>
<td>5</td>
</tr>
<tr>
<td>Act comfortably with others.</td>
<td>2</td>
</tr>
<tr>
<td>Cheer people up.</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Friendliness</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make friends easily.</td>
<td>5</td>
</tr>
<tr>
<td>Warm up quickly to others.</td>
<td>5</td>
</tr>
<tr>
<td>Feel comfortable around people.</td>
<td>5</td>
</tr>
<tr>
<td>Act comfortably with others.</td>
<td>4</td>
</tr>
<tr>
<td>Cheer people up.</td>
<td>5</td>
</tr>
<tr>
<td>Am hard to get to know. (R)</td>
<td>5</td>
</tr>
<tr>
<td>Often feel uncomfortable around others. (R)</td>
<td>4</td>
</tr>
</tbody>
</table>

How likely are these responses given what we know about the respondent and the rest of the sample?
### The response process model

<table>
<thead>
<tr>
<th>Component</th>
<th>Specific Processes</th>
</tr>
</thead>
</table>
| Comprehension | Attend to questions and instructions  
|              | Represent logical form of question  
|              | Identify question focus (information sought)  
|              | Link key terms to relevant concepts |
| Retrieval   | Generate retrieval strategy and cues  
|              | Retrieve specific, generic memories  
|              | Fill in missing details |
| Judgment    | Assess completeness and relevance of memories  
|              | Draw inferences based on accessibility  
|              | Integrate material retrieved  
|              | Make estimate based on partial retrieval |
| Response    | Map judgment onto response category  
|              | Edit response |

Outline

• CR defined

• Importance of CR
  - How frequently does this occur?
  - What impact do these responses have on data?

• Causes of CR

• Ways to Address CR
  – CR Indicators
  – CR Prevention
Prevalence

- Best guesses are that around 10% of cases may be careless on average, though this may fluctuate.
  - 3.5% (Johnson, 2005); 5% (Ehlers et al., 2009); 5%-20% (Curran et al., 2010); 10% (Kurtz & Parish, 2001); 11% (Meade & Craig, 2012)

- Truly random is not very common – it takes considerable cognitive resources
- Extremely consistency is not very common – it stands out

- Most carelessness seems to be intermittent
  - Self-reported occasionally careless rates > 50% (Baer et al., 1997; Berry et al., 1992)
Importance of Careless Responding

\[ X = T + E \]

- Scale score (\(X\)) is a function of true score (\(T\)) plus random error (\(E\))
- Careless responding is a source of error, but it may be random or systematic
Importance of Careless Responding

- Random error will attenuate (lower) correlations
  - Most careless response appears to be non-systematic
- Systematic error may inflate or attenuate correlations

- Attenuated correlations will impact hypothesis tests and factor analyses
  - Careless subsamples show almost no factor structure (Johnson, 2005; Meade & Craig, 2012)
Huang et al. (2012)

- Removal of individuals generally improved factor loadings and produced a cleaner scree plot

- This largely reflects the core of these cases being noise and error

Fig. 1 Scree plot from exploratory factor analysis for vulnerability facet before and after removal of IER detected using page time 99% specificity cutoff. Full sample: before removal of IE responses (N = 345). Trimmed sample: after removal of IE responses (N = 327)
Importance of Careless Responding

- Simulation study illustrated that as little as 10% of careless responding can give rise to positive and negative worded factors (Woods, 2006)
  - Both careful and careless respondents will rate *positively* worded items the same way
  - Both careful and careless respondents will rate *negatively* worded items the same way

<table>
<thead>
<tr>
<th>Friendliness</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make friends easily.</td>
<td>5</td>
</tr>
<tr>
<td>Warm up quickly to others.</td>
<td>5</td>
</tr>
<tr>
<td>Feel comfortable around people.</td>
<td>5</td>
</tr>
<tr>
<td>Act comfortably with others.</td>
<td>4</td>
</tr>
<tr>
<td>Cheer people up.</td>
<td>5</td>
</tr>
<tr>
<td>Am hard to get to know. (R)</td>
<td>5</td>
</tr>
<tr>
<td>Often feel uncomfortable around others. (R)</td>
<td>4</td>
</tr>
</tbody>
</table>
Importance of Careless Responding

- Careful respondents will show high positive correlations between items once reverse recoded.
- Careless respondents will show high negative correlations between items once reverse coded.
- Leads to weak correlations between positive and negative worded items across the sample.

<table>
<thead>
<tr>
<th></th>
<th>+ worded items</th>
<th>- Warded Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ worded items</td>
<td>Strong r</td>
<td></td>
</tr>
<tr>
<td>- worded items</td>
<td>Weak r</td>
<td>Strong r</td>
</tr>
</tbody>
</table>
Importance of Careless Responding

- Another simulation study examined how different patterns of response would impact coefficient alpha (Curran & Kotrba, 2012)
- Lots of consistency in responses can inflate coefficient alpha
- Random responding deflates coefficient alpha
Importance of Careless Responding

• Careless data can do more than simply introduce random noise that attenuates correlations
• Simulation indicates how careless response can actually create spurious results (Huang, Liu, & Bowling, 2015)
Importance of Careless Responding

• Replication
  – If having different types of careless responses in a dataset can impact results, then having differing proportions of these different types of individuals may drive or suppress results in one sample but not another

  – It is important to detail the steps that are taken to clean data
Outline

• CR defined
• Importance of CR
• Causes of CR
• Ways to Address CR
  – CR Indicators
  – CR Prevention
Outline of potential causes

- Survey response involves allocation of resources (cognition, time, energy, etc.). Some relevant theories:
  - Trait theory
  - State causes
    - Distractions
    - Norms
    - Motivation
      - Equity theory
      - Expectancy theory (VIE)
    - Ego depletion
      - Survey length
### Big 5 Correlations

<table>
<thead>
<tr>
<th></th>
<th>O</th>
<th>C</th>
<th>E</th>
<th>A</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Diligence</td>
<td>0.12*</td>
<td>0.15*</td>
<td>0.01</td>
<td>0.16*</td>
<td>-0.02</td>
</tr>
<tr>
<td>Interest</td>
<td>0.14*</td>
<td>0.09*</td>
<td>0.01</td>
<td>0.15*</td>
<td>0.05</td>
</tr>
<tr>
<td>SR Effort</td>
<td>0.10*</td>
<td>0.13*</td>
<td>0.04</td>
<td>0.15*</td>
<td>0.00</td>
</tr>
<tr>
<td>SR Attention</td>
<td>0.15*</td>
<td>0.19*</td>
<td>-0.05</td>
<td>0.13*</td>
<td>0.01</td>
</tr>
<tr>
<td>SRSI UseMe</td>
<td>0.08*</td>
<td>0.11*</td>
<td>0.00</td>
<td>0.07*</td>
<td>-0.03</td>
</tr>
<tr>
<td>Sum of Careless Flags</td>
<td>-0.02</td>
<td>-0.06</td>
<td>0.04</td>
<td>-0.07*</td>
<td>0.00</td>
</tr>
<tr>
<td>Mahalanobis D</td>
<td>-0.03</td>
<td>-0.10*</td>
<td>-0.01</td>
<td>-0.08*</td>
<td>0.12*</td>
</tr>
<tr>
<td>Even-Odd Consistency</td>
<td>0.15*</td>
<td>0.12*</td>
<td>0.04</td>
<td>0.22*</td>
<td>-0.11*</td>
</tr>
<tr>
<td>Psychometric Synonymms</td>
<td>0.14*</td>
<td>0.17*</td>
<td>0.10*</td>
<td>0.17*</td>
<td>-0.11*</td>
</tr>
<tr>
<td>Psychometric Antonymns</td>
<td>0.20*</td>
<td>0.11*</td>
<td>0.03</td>
<td>0.17*</td>
<td>-0.08*</td>
</tr>
<tr>
<td>LongString</td>
<td>-0.01</td>
<td>-0.04</td>
<td>0.02</td>
<td>0.03</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Strongest:**

- Agreeableness
- Openness
- Conscientiousness
Strongest:
- Agreeableness
- Extraversion
- Some facets of conscientiousness

Moderate:
- Openness
### State vs Trait

Everything except openness (Bowling et al., 2016)

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquaintance-reported personality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Agreeableness</td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Conscientiousness</td>
<td>.46</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Emotional stability</td>
<td>.53</td>
<td>.42</td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Openness</td>
<td>.28</td>
<td>.25</td>
<td>.21</td>
<td>.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Extraversion</td>
<td>.16</td>
<td>.29</td>
<td>.35</td>
<td>.31</td>
<td>.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-report questionnaire IER Indices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Individual reliability IER</td>
<td>-.18</td>
<td>-.22</td>
<td>-.21</td>
<td>-.15</td>
<td>-.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Psychological synonyms IER</td>
<td>-.15</td>
<td>-.10</td>
<td>-.22</td>
<td>.04</td>
<td>-.07</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Psychological antonyms IER</td>
<td>-.11</td>
<td>-.18</td>
<td>-.21</td>
<td>.03</td>
<td>-.17</td>
<td>.38</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Mahalanobis D IER</td>
<td>-.20</td>
<td>-.22</td>
<td>-.29</td>
<td>.00</td>
<td>-.20</td>
<td>.79</td>
<td>.80</td>
<td>.77</td>
<td>.75</td>
<td>.79</td>
</tr>
<tr>
<td>10. IER (standardized average)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. IER = insufficient effort responding. N = 215 to 217. αs reported on the diagonal (in italicized font) when applicable. When |r| ≥ .14, p < .05; when |r| ≥ .18, p < .01; when |r| > .22, p < .001.
Respondents completing survey were asked to list features of environment.

Biggest predictors of various CR indicators were reading and sending text messages, IMs, and emails (Osgood, Ward, & Meade, 2015)

<table>
<thead>
<tr>
<th>Distraction</th>
<th>Freq.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>I could see people or things moving in the background.</td>
<td>82</td>
<td>36.8</td>
</tr>
<tr>
<td>I received an alert indicating an incoming email, text message, instant message, online chat request, or phone call.</td>
<td>81</td>
<td>36.3</td>
</tr>
<tr>
<td>Someone spoke to me in person.</td>
<td>71</td>
<td>31.8</td>
</tr>
<tr>
<td>I was listening to music.</td>
<td>67</td>
<td>30.0</td>
</tr>
<tr>
<td>I read a text messages.</td>
<td>64</td>
<td>28.7</td>
</tr>
<tr>
<td>I sent a text message.</td>
<td>57</td>
<td>25.6</td>
</tr>
<tr>
<td>There was a television active in the area.</td>
<td>46</td>
<td>20.6</td>
</tr>
<tr>
<td>I looked at another webpage other than the survey page.</td>
<td>33</td>
<td>14.8</td>
</tr>
<tr>
<td>I attended to music (e.g. skipping a song, changing the volume, or pausing/resuming play).</td>
<td>25</td>
<td>11.2</td>
</tr>
<tr>
<td>My cell/home phone rang or vibrated to indicate an incoming call.</td>
<td>18</td>
<td>8.1</td>
</tr>
<tr>
<td>I read an e-mail.</td>
<td>11</td>
<td>4.9</td>
</tr>
<tr>
<td>I read an instant message.</td>
<td>10</td>
<td>4.5</td>
</tr>
<tr>
<td>I spoke with someone on my cell/home phone.</td>
<td>9</td>
<td>4.0</td>
</tr>
<tr>
<td>I sent an instant message.</td>
<td>7</td>
<td>3.1</td>
</tr>
<tr>
<td>I sent an e-mail.</td>
<td>3</td>
<td>1.3</td>
</tr>
<tr>
<td>I made a post on a webpage (such as Facebook or internet forum).</td>
<td>2</td>
<td>0.9</td>
</tr>
</tbody>
</table>
State Causes: Norms

• Dillman and others argue that the survey response process is a type of social interaction

• Unproctored surveys typically lack any interaction between the researcher and respondent

• Lower incidence of careless responding in person than online (Bowling et al., in press; Francavilla et al., in press)
State Causes: Motivation theories

- Equity theory: If respondents feel their reward to input ratio is lower than it should be they will be motivated to
  - Increase rewards
  - Reduce effort

- Expectancy theory: Respondents will be motivated if they
  - Desire a promised outcome
  - Believe they can complete the survey
  - Trust that the outcome will be delivered as promised
Ego Depletion

- Attention is a resource, thus it is likely that as survey length increases, carelessness will increase.
- It is possible that carelessness serves as a “break” during which respondents attempt to regain attentional resources.
- Survey length is associated with an increase in careless response (Meade & Craig, 2012).
- In a randomized design study, longer surveys were associated with large increase in items skipped and small increase in careless response (Bowling et al., in press).
Outline

• CR defined
• Importance of CR
• Causes of CR
• **Ways to Address CR**
  – CR Indicators
  – CR Prevention
Indicators overview

• Consistency indices
  – Even-odd correlations
  – Semantic and Psychometric synonyms/antonyms correlations
  – LongString
• Outlier analysis: Mahalanobis Distance
• Other
  – Response Time
  – Self-report
  – Special items and instructions
Planning Ahead vs Post-Hoc

- Methods that require planning
  - Response time
  - Bogus/infrequency items
  - Attention check items
  - Instructional manipulation checks
  - Self-report scales
- Post-hoc methods
  - Most consistency methods
  - Outlier analysis

- Note: These indicators behave more like clinical measures than trait measures. Most require some type of cutoff value.
Response Time

- The amount of time an individual takes to respond to all items (per page, per survey, etc.)
- Recommended cutoff is 2 seconds per item for page time (Huang et al., 2012)

- Only the low end of the scale is examined

- **Pros**: One of the most straightforward metrics
- **Cons**: Only effective at detecting the most obvious respondents; overall time misses distracted respondents
LongString Analysis

- One of the ways that individuals can respond carelessly is by providing the same response to every question.
- LongString is the maximum number of times the same response is used consecutively.
- Range of values: 1 to k.

**Pros:** Effective for overly consistent respondents.

**Cons:** Only detects one type of carelessness; cutoff needed.
Outlier: Mahalanobis Distance

- Multivariate outlier looking at squared residual from principal component space
- Similar to sum of a squared z-score that takes into account the relationships among the variables

Once I've made up my mind, other people can seldom change my opinion.

I never hesitate to go out of my way to help someone in trouble.

I use flattery to get ahead.
Outlier: Mahalanobis Distance

• **Pros**
  – You can compute a chi-square value for Mahalanobis D which provides cutoff values
  – Can detect more complex types of carelessness
  – Some evidence of efficacy (Meade & Craig, 2012)

• **Cons**
  – May miss some types of careless responding (e.g., mid-point carelessness)
Consistency: Odd-Even Correlation

- **Theory**: Items that measure the same thing should produce similar response
- Within-person correlation

- Assuming scales have sufficient unidimensionality, scores computed on halves of each scale should be related to each other
- Each scale is broken into odd and even items, and scores are computed for those halves for each respondent
- These halves are paired, and correlations are computed across multiple scales for each respondent
Consistency: Even-Odd Correlation

- Scale A: i1, i2, i3, i4, i5, i6
- Scale B: i7, i8, i9, i10, i11, i12, i13, i14
- ...
- Scale Z: i99, i100, i101, i102, i103, i104

<table>
<thead>
<tr>
<th>Scale A_{odd}</th>
<th>Scale A_{even}</th>
</tr>
</thead>
<tbody>
<tr>
<td>((i1 + i3 + i5)/3)</td>
<td>((i2 + i4 + i6)/3)</td>
</tr>
<tr>
<td>Scale B_{odd}</td>
<td>Scale B_{odd}</td>
</tr>
<tr>
<td>((i7 + i9 + i11 + i13)/4)</td>
<td>((i8 + i10 + i12 + i14)/4)</td>
</tr>
</tbody>
</table>
| ... | ...
| Scale Z_{odd} | Scale B_{odd} |
| \((i99 + i101 + i103)/3\) | \((i100 + i102 + i104)/3\) |

Correlate the two columns to get even-odd correlation
Consistency: Odd-Even Correlation

• Range of values: -1 to 1
• Lower values = more carelessness

• **Pros**: Post-hoc, logical, some evidence of efficacy (Meade & Craig, 2012)

• **Cons**: Depends on unidimensional scales; reliability of index can be poor with few scales
Consistency: Resampled Internal Reliability

• Odd-Even Consistency is a special case of split halves
• There are many other potential halves that can be used
• Resampling these halves and computing the mean of these correlations should give a better approximation than simply using odd-even halves
• **Pros**
  – Same as even-odd consistency but less prone to sampling error
• **Cons**
  – Same as even-odd consistency plus more difficult to compute
Consistency: Semantic/Psychometric Synonyms

- **Theory**: Items that measure the same thing should produce similar response
- **Within-person correlation**
- **Item pairs** can be chosen a priori, based on semantic meaning of items (semantic synonyms) or after the fact based on sample correlations (psychometric synonyms)
- Same thing can be done for item pairs where opposite responses are expected (e.g., reverse coded items). This is the semantic/psychometric **antonym** approach
- **When using correlations**, sometimes item pairs can come from different scales but must choose correlation cutoff (e.g., .60; Meade & Craig, 2012)
Consistency: Semantic/Psychometric
Synonyms/Antonyms

- Pairs of items are selected:
  - \((x_1, x_2), (y_1, y_2), (z_1, z_2)\)

- The number of these pairs is the “sample size” of the resulting correlation between the two vectors:

\[
\begin{bmatrix}
  x_1, & y_1, & z_1 \\
  x_2, & y_2, & z_2 \\
\end{bmatrix}
\]
Consistency: Semantic/Psychometric Synonyms/Antonyms

- Range of values: -1 to 1
- Lower values = more carelessness
- **Pros**: Post-hoc, logical, works reasonably well (Meade & Craig, 2012), does not rely on existing scales
- **Cons**: must set cutoff for both correlation needed to form a pair and overall, carelessness in sample affects correlations among pairs
Consistency: Person-Total Correlation

- Person-total correlation is a flip of item-total correlation
- Measures consistency compared to other respondents
Consistency: Person-Total Correlation

- Range of values: -1 to 1
- Lower values = more carelessness

- **Pros**: Post-hoc, logical, slightly different from other indices, doesn’t require item pairs or scales
- **Cons**: less research on this index
Check Items

- Normal survey items don’t have ‘correct’ answers, they only have valid answers; we don’t know true values to check this
- We can insert items with ‘correct’ choices
- Individuals who endorse these items incorrectly are flagged as careless responders

- **Instructed response items:** Please choose ‘strongly disagree’ for this item.
- **Pros:** logical and clear, no cutoff needed
- **Cons:** respondents may be on the alert for these items, does not detect sporadic carelessness well, may annoy respondents
Check Items: Bogus/Infrequency Items

Meade & Craig (2012)

Table 1
Bogus Items Response Rates

<table>
<thead>
<tr>
<th>Item</th>
<th>Str. D1</th>
<th>D2</th>
<th>Sl.D3</th>
<th>Neither A nor D4</th>
<th>Sl. A5</th>
<th>A6</th>
<th>Str. A7</th>
<th>% Flagged by item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I am using a computer currently. (R)</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>55</td>
<td>315</td>
<td>4</td>
</tr>
<tr>
<td>2. I have been to every country in the world.</td>
<td>299</td>
<td>59</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>3. I am enrolled in a Psychology course currently. (R)</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>53</td>
<td>310</td>
<td>7</td>
</tr>
<tr>
<td>4. I have never spoken to anyone who was listening.</td>
<td>116</td>
<td>123</td>
<td>36</td>
<td>90</td>
<td>9</td>
<td>10</td>
<td>2</td>
<td>38*</td>
</tr>
<tr>
<td>5. I sleep less than one hour per night.</td>
<td>286</td>
<td>60</td>
<td>13</td>
<td>15</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>6. I do not understand a word of English.</td>
<td>301</td>
<td>48</td>
<td>7</td>
<td>12</td>
<td>5</td>
<td>11</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>7. I have never brushed my teeth.</td>
<td>322</td>
<td>33</td>
<td>6</td>
<td>12</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>8. I am paid biweekly by leprechauns.</td>
<td>269</td>
<td>38</td>
<td>13</td>
<td>43</td>
<td>8</td>
<td>4</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>9. All my friends are aliens.</td>
<td>270</td>
<td>47</td>
<td>15</td>
<td>28</td>
<td>8</td>
<td>7</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>10. All my friends say I would make a great poodle.</td>
<td>215</td>
<td>66</td>
<td>21</td>
<td>62</td>
<td>8</td>
<td>11</td>
<td>3</td>
<td>27</td>
</tr>
</tbody>
</table>

*Note.* D = disagree; A = agree; Str. = strongly; Sl. = slightly; R = Items flagged if (reverse coded) strongly disagree or disagree not chosen (except missing data).

*Item 4 was dropped as a Bogus Item based on frequent response.*
Check Items: Bogus/Infrequency Items

Huang, Bowling, Liu, & Li (2014)

Table 1  Confirmatory factor analysis for the eight-item IER Scale for Study 1

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I can run 2 miles in 2 min</td>
<td>0.16</td>
<td>0.36</td>
<td>0.81***</td>
</tr>
<tr>
<td>2. I eat cement occasionally</td>
<td>0.11</td>
<td>0.31</td>
<td>0.97***</td>
</tr>
<tr>
<td>3. I can teleport across time and space</td>
<td>0.16</td>
<td>0.37</td>
<td>0.93***</td>
</tr>
<tr>
<td>4. I am interested in pursuing a degree in parabanjology</td>
<td>0.16</td>
<td>0.36</td>
<td>0.87***</td>
</tr>
<tr>
<td>5. I have never used a computer</td>
<td>0.10</td>
<td>0.29</td>
<td>0.96***</td>
</tr>
<tr>
<td>6. I work fourteen months in a year</td>
<td>0.15</td>
<td>0.35</td>
<td>0.82***</td>
</tr>
<tr>
<td>7. I will be punished for meeting the requirements of my job</td>
<td>0.17</td>
<td>0.37</td>
<td>0.80***</td>
</tr>
<tr>
<td>8. I work twenty-eight hours in a typical work day</td>
<td>0.02</td>
<td>0.13</td>
<td>0.44*</td>
</tr>
</tbody>
</table>

N = 284

* p < 0.05; *** p < 0.001. Standardized loadings are shown
Check Items: Bogus/Infrequency/Check Items

- Range of values: 0 to count of items
- More incorrect endorsements = more careless

- **Pros:** Blend well with other survey items
- **Cons:**
  - More subjective than instructed response
  - Respondents may unexpectedly agree
  - Must set cutoff
Check Items: Instructional Manipulation Checks

- This technique is most commonly used in situations where important information or manipulation is given in the study instructions.

Fig. 3. Example of an IMC based on Oppenheimer et al. (2009).
Check Items: Instructional Manipulation Checks

- Range of values: correct or incorrect
- Incorrect answers = more carelessness
- **Pros**
  - Logical
  - No need for cutoff values
- **Cons**
  - Space consuming
  - Does not necessarily indicate carelessness on items responses
Self-Report: Scales and Questions

Perhaps the easiest way to screen is just to ask the respondent
Lastly, it is vital to our study that we only include responses from people that devoted their full attention to this study. Otherwise years of effort (the researchers and the time of other participants) could be wasted. You will receive credit for this study no matter what, however, please tell us how much effort you put forth towards this study:

I put forth ______ effort towards this study.

Also, often there are several distractions present during studies (other people, TV, music, etc). Please indicate how much attention you paid to this study. Again, you will receive credit no matter what. We appreciate your honesty!

I gave this study ____ attention

In your honest opinion, should we use your data in our analyses in this study? Yes/No

From Meade & Craig (2012)
## Self-Report: Diligence & Interest

**Table A1**

*Exploratory Factor Analysis of Participant Engagement Items*

<table>
<thead>
<tr>
<th>Item</th>
<th>Diligence</th>
<th>Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I carefully read every survey item.</td>
<td>.71</td>
<td>.01</td>
</tr>
<tr>
<td>2. I could’ve paid closer attention to the items than I did.</td>
<td>.71</td>
<td>−.16</td>
</tr>
<tr>
<td>3. I probably should have been more careful during this survey.</td>
<td>.71</td>
<td>−.08</td>
</tr>
<tr>
<td>4. I worked to the best of my abilities in this study.</td>
<td>.71</td>
<td>.12</td>
</tr>
<tr>
<td>5. I put forth my best effort in responding to this survey.</td>
<td>.65</td>
<td>.05</td>
</tr>
<tr>
<td>6. I didn’t give this survey the time it deserved.</td>
<td>.59</td>
<td>−.04</td>
</tr>
<tr>
<td>7. I was dishonest on some items.</td>
<td>.55</td>
<td>−.05</td>
</tr>
<tr>
<td>8. I was actively involved in this study.</td>
<td>.48</td>
<td>.19</td>
</tr>
<tr>
<td>9. I rushed through this survey.</td>
<td>.45</td>
<td>.15</td>
</tr>
<tr>
<td>10. I enjoyed participating in this study.</td>
<td>−.02</td>
<td>.82</td>
</tr>
<tr>
<td>11. This study was a good use of my time.</td>
<td>−.16</td>
<td>.71</td>
</tr>
<tr>
<td>12. I was bored during the study.</td>
<td>−.04</td>
<td>.65</td>
</tr>
<tr>
<td>13. This survey was too long.</td>
<td>−.06</td>
<td>.59</td>
</tr>
<tr>
<td>14. The work I did for this study is important to me.</td>
<td>.09</td>
<td>.57</td>
</tr>
<tr>
<td>15. I care about my performance in this study.</td>
<td>.29</td>
<td>.50</td>
</tr>
<tr>
<td>16. I would be interested in reading about the results of this study.</td>
<td>.18</td>
<td>.38</td>
</tr>
<tr>
<td>17. I’m in a hurry right now.</td>
<td>.20</td>
<td>.29</td>
</tr>
</tbody>
</table>

*Note.* Items 16 and 17 were not retained. Bold indicates loadings > .4.

From Meade and Craig, 2012
Self-Report

• **Pros**
  – Easy to incorporate
  – Transparent

• **Cons**
  – Low correlation with other indicators
  – Prone to social desirability
  – Some need cutoffs
General Recommendations

- Should be routine part of data cleaning process
- There is no one method which is better than any other in all cases so choose from the following categories
  - LongString
  - 1 or 2 consistency indices based on within-person correlation
  - Outlier (e.g., Mahalanobis)
  - Also consider: Time and self-report
- Treated indicators separately – do NOT combine into a single index
  - Consider multiple hurdle or other non-compensatory approach
- Use empirical distributions to make informed choices for cutoff values
- Be mindful of the percentage of your sample you are removing
- Document what you have done
- R package: https://github.com/ryentes/careless
Outline

• CR defined
• Importance of CR
• Causes of CR
• Ways to Address CR
  – CR Indicators
  – CR Prevention
Prevention-focused Studies

• Survey instructions influence CR
  – Identified responses
  – Warning
  – Increasing perceived social interaction
  – Cognitive Dissonance (Hypocrisy)
  – Reciprocity

• Survey design
  – Virtual human
Prevention-focused Studies

• Survey instructions influence CR
  – Identified responses
    • Significant but small effect of identified responses (Meade & Craig; 2012)
  – Warning
    • Significant small/medium effect of warnings (Meade & Craig; 2012; Huang et al., 2012, Ward, 2013)
  – Increasing perceived social interaction
    • Providing a video introduction to the study had minimal effect (Meade & Ward, 2014)
Prevention-focused Studies

- Survey instructions influence CR
  - Cognitive Dissonance (Hypocrisy)
    - Small but significant effects of asking participants to list reasons to pay attention and to promise not to be careless (Ward et al., 2015)
  - Reciprocity
    - Small but significant effects of telling respondents that a donation had been made in their name or that researchers had obtained free snack for them from a local business (Ward et al., 2016)
- Survey design
  - Virtual human
    - Virtual proctor not as effective as in-person proctors but can have a small effect (Francavilla & Meade, 2016; Ward, 2014)
Summary of Findings on Prevention

- Warning reduces CR a little
- Sometimes self-report does not match behavioral indices
- Increasing perceived social interaction doesn’t reduce objective indicators of CR
- Cognitive dissonance and hypocrisy reduces some forms of CR
- Virtual human in combination with warning instructions reduces CR a little