Evaluating and Improving Quality of Measurement

Moderator: Mary McDougall
CEO, CfMC Survey Management Solutions

Friday, May 16, 2014

Re-presented at PAPOR Mini-Conference
June 13, 2014
San Francisco, CA
Disclaimer:

Slides were obtained with author permission but any mistakes/misinterpretations are my own!

Please contact the authors if you have additional questions about their work.
Quality of Measurement

- Comparability of Common Items Across Survey and Administrative Datasets
- Making Use of Proxy Reports in a Telephone Survey
- Characterization and Analysis of Duplicate Responses in a Physician Survey
- Decreasing “Satisficing” in Web-Surveys – Evidence from an Awareness Control Experiment
- Reconciling Public Participation Rate Differences in Census Bureau vs. Academic/Commercial Survey Estimates
Comparability of Common Items Across Survey and Administrative Datasets

Does a “no” to a question in one survey mean the same thing as a “no” to a similar question in a different survey?

This presentation shows that the interpretations of similar questions and answer categories across different survey environments can be widely divergent.
Comparability of Common Items across Survey and Administrative Datasets

Paul J Scanlon, PhD
National Center for Health Statistics (NCHS)
Overview

- A “no” to one survey question doesn’t necessarily mean a “no” to another similar question on another survey or form...

- Traditional cognitive interviewing does not find these differences, but qualitative validity tests can.

- Using validity tests, we can compare the interpretations of similar items across instruments to determine whether or not the data should be linked.
Validity Across Datasets

• New Landscape: Increased importance of “big data” and administrative data/survey linkage

• Context matters to respondents

• Validity testing is the pragmatic & cost-effective way to move forward
A Shift in Cognitive Interviewing...

• Traditionally used to find survey methodology and cognitive “problems.”

• Transition to elucidating “interpretations” instead of “problems.”

• We can compare constructs across survey and administrative instruments.
Example: Is hearing a conversation in a loud room a problem for you?

- Traditional Cognitive Interviewing Findings:
  - No comprehension issues
  - Respondents understand “loud room” to be parties, restaurants, bars, etc

- Qualitative Validity Testing Findings:
  - Rs who answered in the affirmative tended to be thinking about their physical ability to hear—**their health conditions**
  - Rs who answered in the negative tended to be thinking about the fact that it’s **always more difficult to hear in a loud room**—that’s what makes it loud.
By “No” I mean “Don’t Know”...

- Validity testing allows us to look across groups within a survey population

- NIOSH Survey of Respiratory Health of Healthcare workers

  Asked about presence of specific chemicals
By “No” I mean “Don’t Know”...

- Validity testing allows us to look across groups within a survey population
- NIOSH Survey of Respiratory Health of Healthcare workers
  - Testing on both supervisors and line workers
  - Interpretation of “Chemical Use” questions varied by job level.
  - For line workers, “no” often meant “don’t know,” while for managers “no” almost always meant “no.”

Difference between groups was related more to EXPERIENCE than EDUCATION
Is there a solution?

- No magic wand

- Increased use of validity testing

- Look across validity tests for trends
  - Experience levels in establishment surveys and forms
  - Education level in household surveys and forms
Characterization and Analysis of Duplicate Responses in a Physician Survey

This presentation looks at issues related to the relative quality and quantity of information provided by proxies based on a survey by National Center for Health Statistics.
Characterization & Analysis of Duplicate Responses in a Physician Survey

Julianne Payne, PhD¹; Julie Linville, MA¹; Paul Beatty, PhD²; Nicholas A. Holt, PhD¹

¹Social & Scientific Systems, Inc.; ²National Center for Health Statistics

American Association of Public Opinion Research
May 16, 2014
Research Questions

Does target and proxy response differ? If so, how?

What are the implications of using proxy responses in place of target responses?
Data

- 2011-2013 National Electronic Health Records Survey (NEHRS)
- Examine “duplicates” – N=1,034 surveys for 517 physicians
  - N=343 physician submitted twice
  - N=49 staff submitted twice
  - N=125 respondent varied

Target = physician; Proxy = office staff
### Results: Item Missingness

<table>
<thead>
<tr>
<th>Domain</th>
<th>% Missing, Target</th>
<th>% Missing, Proxy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey Focus</td>
<td>23.9</td>
<td>7.4</td>
</tr>
<tr>
<td>Org Structure</td>
<td>11.5</td>
<td>4.3</td>
</tr>
<tr>
<td>Administration</td>
<td>23.2</td>
<td>6.9</td>
</tr>
<tr>
<td>Clinical Treatment</td>
<td>23.1</td>
<td>6.9</td>
</tr>
</tbody>
</table>

“Does your practice use electronic medical records/electronic health records?”

13.2% missing among physicians

3.1% missing among proxies

Proxies answered more completely
### Results: Content of Response, Study Focus

<table>
<thead>
<tr>
<th>Survey Question</th>
<th>% of Matching Responses, Missing as Valid</th>
<th>% of Matching Responses, Missing as Invalid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Both Target</td>
<td>Both Proxy &amp; Proxy</td>
</tr>
<tr>
<td>Practice uses EHR</td>
<td>84.8</td>
<td>83.7</td>
</tr>
<tr>
<td>Name of EHR</td>
<td>89.2</td>
<td>85.7</td>
</tr>
<tr>
<td>Year EHR installed</td>
<td>75.2</td>
<td>79.6</td>
</tr>
<tr>
<td>Plans to install EHR</td>
<td>79.0</td>
<td>67.3</td>
</tr>
</tbody>
</table>

On Focus: Proxies and Targets gave different answers

20% disagreement between targets and proxies on use of EHR –

77% physicians report use, vs. 64% proxies
Results: Content of Response, Non-focal Items

Percent of Matching Responses by Duplicate Type, Missing as Valid

- Org Structure: Both Target 89, Both Proxy 87, Target & Proxy 79
- Administration: Both Target 83, Both Proxy 76, Target & Proxy 66
- Clinical Treatment: Both Target 80, Both Proxy 73, Target & Proxy 60

Note: Answers even differed within Respondent types
Conclusions

• Little variation in survey completion by participant type
• Proxy surveys contain fewer missing responses
• Significant within-participant disagreement across responses
• Greater across-participant disagreement
Lessons for the NEHRS

• High within-type variation raises questions
  • Difficulty of survey questions
  • Participant motivation
  • Speed of organizational change

• Examine barriers to physician response

• Consider alternatives to proxies

• Analysts should be deliberate about selecting from duplicates for analytic purposes
Making Use of Proxy Reports in a Telephone Survey

This presentation discusses the pros and cons of using proxy respondents in a recent telephone survey to determine:

*The feasibility of collecting national estimates of the prevalence of educational and work-related credentials among the U.S adult population.*
Making Use of Proxy Reports in a Telephone Survey

Rebecca Medway and Celeste Stone
Methods: Research Questions

1. **How many** people were willing to provide a proxy report?
   - Proportion of respondents in proxy experiment willing to provide proxy report

2. **Who** is willing to provide a proxy report? **Who** are proxies reporting on?
   - Demographic characteristics

3. **How well** do proxy respondents report on our variables of interest?
   - Extent of DK responses
   - Validity of reports
Data: Proxy Experiment

- Half of sampled adults living in household with two or more people were randomly selected for proxy experiment
- After completing self report, they were asked to complete a second interview about a pre-identified proxy subject (another adult household member)

2+ ppl in HH; randomly selected for exp.
Results:
Who – Proxy Reporters

- Among those respondents asked to complete a second, proxy interview, those who completed one were significantly younger on average than those who did not.

### Reporter Demographics by Willingness to Provide Proxy Report

<table>
<thead>
<tr>
<th>Gender</th>
<th>Completed proxy report (n=1040)</th>
<th>Did not end up completing proxy report (n=120)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>46%</td>
<td>43%</td>
</tr>
<tr>
<td>At least some college</td>
<td>68%</td>
<td>59%</td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>82%</td>
<td>78%</td>
</tr>
</tbody>
</table>

### Reporter Mean Age by Willingness to Provide Proxy Report

<table>
<thead>
<tr>
<th>Interview type</th>
<th>Mean age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed proxy report (n=1040)</td>
<td>50.7</td>
</tr>
<tr>
<td>Did not complete proxy report (n=120)</td>
<td>54.2</td>
</tr>
</tbody>
</table>
Results:
Who – Subject of Proxy Reports

- The individuals who provided the proxy reports were significantly more likely than the people they were reporting about to have completed at least some college.
Results: How Well – DK Response to Main Credential Item

- Proxy reporters are more likely to provide DK responses for intro question asking whether or not respondent has a work credential.

DK Responses by Report Type: Main Credential Item

- 100%
- 75%
- 50%
- 25%
- 1%
- 7%
- 0%

Self report (n=1040)  Proxy report (n=1040)

True regardless of age or education

Significant difference between self and proxy (p<0.05)
Results:
How Well – Validity of Credential Reports

- Significantly fewer credentials in proxy reports than in self reports.
- No significant differences between proxy responses and self-reported responses for the key items used to estimate overreporting (excluding DK responses).

Responses to Overreporting Items:

- Have credential: 35% (Self Report) vs 30% (Proxy Report)
- Personal-interest: 5% (Self Report) vs 7% (Proxy Report)
- No assessment: 5% (Self Report) vs 3% (Proxy Report)
- Not portable: 3% (Self Report) vs 2% (Proxy Report)

Significant difference between self and proxy (p<0.05)
Conclusions

- **How many**
  - Most people asked to conduct a proxy interview ended up doing so

- **Who**
  - Younger people were more likely to provide a proxy report
  - Allowing proxy reports may bring in reports about harder-to-reach groups (less educated)

- **How well**
  - Proxy reporters less likely to know the answers to the survey questions (DK responses).
  - However, when credentials were reported, proxy reporters were no more likely than self reporters to overreport for key estimates.

- **Future research:**
  - Relationship between proxy reporter and subject effect quality of proxy report?
  - Higher response rate survey result in lower proxy cooperation rate?
Reconciling Public Participation Rate Differences in Census Bureau vs. Academic/Commercial Survey Estimates

Relatively small differences in Census Bureau figures can result in dramatic changes in public policy and planning. When the Survey of Public Participation in the Arts was established, it was conducted by the Census Bureau to provide the most accurate possible benchmarks of participation in the arts in the context of other free-time activities. For each arts activity, however, the Census figures have been significantly lower than those from Westat and other respected independent surveys.

*This presentation explores several explanations for these lower Census Bureau participation rates.*
Evaluating and Improving Quality of Measurement

Reconciling Public Participation Rate Differences in Census Bureau Vs. Academic/Commercial Survey Estimates

John Robinson, Sociology, University of Maryland
Tim Triplett, Senior Survey Methodologist, Urban Institute

Friday, May 16, 2014
Agenda

• New Census Bureau Undercounts
• Large and consistent difference between Census and Commercial/Academic estimates
Challenge

• Although the match between arts participation questions and years of study is not exact, three separate survey organizations using different field procedures and representing different types of survey organizations have all produced consistently higher estimates of arts and leisure participation than the SPPA surveys conducted by the Census Bureau.

• The Challenge is to try to explain why there are these differences
Comparison of National Surveys Asking Arts Participation Questions

<table>
<thead>
<tr>
<th>Organization</th>
<th>Years</th>
<th>Approximate Sample Size</th>
<th>Mode</th>
<th>Approximate Response Rate</th>
<th>Question Context</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SPPA:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) a) Census Bureau</td>
<td>1982-1985</td>
<td>12,000 – 18,000</td>
<td>¾ Personal ¼ Telephone</td>
<td>90%</td>
<td>Exit questions in Crime Survey</td>
</tr>
<tr>
<td>b) Census Bureau</td>
<td>1992</td>
<td>12,000</td>
<td>¾ Telephone ¼ Personal</td>
<td>70% Jan.–June* 85% July-Dec.</td>
<td>Exit questions in Crime Survey</td>
</tr>
<tr>
<td>c) Census Bureau</td>
<td>2002-2012</td>
<td>17,000</td>
<td>4/5 Telephone 1/5 Personal</td>
<td>66% 2002 75% 2008 72% 2012</td>
<td>Supplement on the Current Population Survey (CPS)**</td>
</tr>
<tr>
<td><strong>2) Westat</strong></td>
<td>1997</td>
<td>12,000</td>
<td>Telephone</td>
<td>55%</td>
<td>Stand-alone Arts Survey</td>
</tr>
<tr>
<td><strong>Non-SPPA:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3) General Social Survey (GSS) (University)</td>
<td>1993-2003</td>
<td>1,500</td>
<td>Personal</td>
<td>70 – 80%</td>
<td>Middle of Omnibus Survey</td>
</tr>
<tr>
<td>4) Univ. MD (University)</td>
<td>1998</td>
<td>1,504</td>
<td>Telephone</td>
<td>57%</td>
<td>Near end of Omnibus Survey</td>
</tr>
</tbody>
</table>
Arts Participation Questions
(Source: Census)

• With the exception of elementary of high school performances, did you go to a live performance between {today’s date} 2011 and {today’s date} 2012?
  – Jazz
  – Classical music such as symphony, chamber or choral music
  – Opera
  – Musical stage play
  – Non-Musical stage play
  – Ballet

• Visited an art museum or Art Gallery last 12 months

• Read literature: last 12 months read any novels, short stories, or plays
Arts Participation: 1982-2012 (excluding 1997)
(Source: Census)
High & Low Census Participation Rates Versus Westat Rates

- Read Literature
- Art Museum
- Musicals
- Plays
- Classical
- Jazz
- Ballet
- Opera

- Census High
- Westat
- Census Low
Summary

- Differences are not due to lower response rates
- Exactly same questions and order in all surveys
- This looks to be a real house effect; possibly due to:
  - Census interviewers better able to locate and interview less active people
  - Census respondents feel less reason to portray themselves as being more active

Differences should be understood before making public policy decisions.
Decreasing “Satisficing” in Web-Surveys – Evidence from an Awareness Control Experiment

Satisficing is a well-known data quality problem in all types of survey research. Contemporary online behavior might, therefore, intervene in the response process and increase the likelihood of satisficing in web surveys compared to other survey modes.

*This presentation discusses how to decrease this likelihood by introducing various “speed controls” or “pauses” in the survey.*
Decreasing Satisficing in Web Surveys
- Evidence from an Awareness Control Experiment

Sebastian Lundmark, Sebastian.Lundmark@gu.se
Johan Martinsson, Johan.Martinsson@pol.gu.se
Department of Political Science, University of Gothenburg, Sweden

Working Paper Presented at American Association for Public Opinion Research’s 69:th Annual Conference,
May 15 – May 18:th, Anaheim, CA, USA.
Awareness Control

Satisficing is a well-known data quality problem in all types of survey research.

In a new digital era the introduction of web surveys has generated great opportunities to control and collect data on how respondents answer question.

However, contemporary online behavior (social media and the likes), might intervene in the response process and increase the likelihood of satisficing in web surveys compared to other survey modes.

This presentation discusses how to possibly decrease this likelihood by introducing various “speed controls” or “pauses” in the survey
Awareness Control

- Instructional Manipulation Checks / Awareness Controls are common when performing experimental surveys studies to ensure that the respondents read the stimuli’s and vignettes.

- However, when performing surveys, we are not only interested in the quality of the stimuli but also in the overall data quality of all the questions we ask.

- We therefore propose that maybe we can use these Instructional Manipulation Checks / Awareness Controls to increase overall data quality amongst the respondents.
Design

• To test this concept we performed a 3x2 full factorial design introducing a new type of Awareness Control as well as asking respondents to Motivate some of their answers.

Awareness Control:

“In the public debate a lot of opinions and proposals exist. In our surveys we often ask about just these sorts of opinions and proposals. These kinds of questions are often followed by a scale by which we ask you to rate to what extent you agree to a certain proposition. The following question serves the purpose to guarantee the quality of our surveys. We therefore ask you to read this whole question thoroughly and, if you read the whole question, answer number two on the following scale.”

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

Strongly Disagree

Strongly Agree
Motivation

• **First:** “On the previous page you answered a question whether you like or dislike political party leaders in the Swedish Parliament. On that question you answered that you generally [like/neither like or dislike/dislike] [party leader name]. We would like to know more about your attitude towards [party leader name]. Please describe, using your own words, the main reasons why you think this way.”

• **Second:** “On the previous page you answered a question on how the Swedish Economy have changed the last 12 months and whether that first and foremost is a product of the government’s politics or mainly other factors. You answered [number] on a scale ranging from 1 to 7, where 1 means that the change is fully from other factors and 7 means that it’s a fully a product of the government’s politics. Please describe, using your own words, the main reasons why you think this way.”
# 1. Results – Failed/Successful of Awareness Control

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>(Correct)</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness Control</td>
<td>0.00</td>
<td>93.23</td>
<td>0.00</td>
<td>0.97</td>
<td>0.97</td>
<td>1.61</td>
<td>3.23</td>
<td>310</td>
</tr>
<tr>
<td>Awareness Control + Force</td>
<td>0.00</td>
<td>93.01</td>
<td>0.91</td>
<td>1.22</td>
<td>0.91</td>
<td>0.91</td>
<td>3.04</td>
<td>329</td>
</tr>
<tr>
<td>Awareness Control + Motivation</td>
<td>0.62</td>
<td>91.36</td>
<td>0.31</td>
<td>1.54</td>
<td>0.62</td>
<td>2.47</td>
<td>3.09</td>
<td>324</td>
</tr>
<tr>
<td>Awareness Control + Force + Motivation</td>
<td>0.32</td>
<td>91.77</td>
<td>0.00</td>
<td>2.53</td>
<td>0.63</td>
<td>2.53</td>
<td>2.22</td>
<td>316</td>
</tr>
<tr>
<td>Total</td>
<td>0.23</td>
<td>92.34</td>
<td>0.31</td>
<td>1.56</td>
<td>0.78</td>
<td>1.88</td>
<td>2.89</td>
<td>1279</td>
</tr>
</tbody>
</table>
1. Results – Failed/Successful of Awareness Control

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>(Correct)</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness Control</td>
<td>0.00</td>
<td>93.23</td>
<td>0.00</td>
<td>0.97</td>
<td>0.97</td>
<td>1.61</td>
<td>3.23</td>
<td>310</td>
</tr>
<tr>
<td>Awareness Control + Force</td>
<td>0.00</td>
<td>93.01</td>
<td>0.91</td>
<td>1.22</td>
<td>0.91</td>
<td>0.91</td>
<td>3.04</td>
<td>329</td>
</tr>
<tr>
<td>Awareness Control + Motivation</td>
<td>0.62</td>
<td>91.36</td>
<td>0.31</td>
<td>1.54</td>
<td>0.62</td>
<td>2.47</td>
<td>3.09</td>
<td>324</td>
</tr>
<tr>
<td>Awareness Control + Force + Motivation</td>
<td>0.32</td>
<td>91.77</td>
<td>0.00</td>
<td>2.53</td>
<td>0.63</td>
<td>2.53</td>
<td>2.22</td>
<td>316</td>
</tr>
<tr>
<td>Total</td>
<td>0.23</td>
<td>92.34</td>
<td>0.31</td>
<td>1.56</td>
<td>0.78</td>
<td>1.88</td>
<td>2.89</td>
<td>1279</td>
</tr>
</tbody>
</table>
1. Awareness Control + Force Follow-up Question

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>(Correct)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness Control + Force</td>
<td>0.00</td>
<td>40.00</td>
<td>8.00</td>
<td>4.00</td>
<td>4.00</td>
<td>12.00</td>
<td>32.00</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness Control + Force + Motivation</td>
<td>3.33</td>
<td>53.33</td>
<td>0.00</td>
<td>13.33</td>
<td>3.33</td>
<td>10.00</td>
<td>16.67</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1.82</strong></td>
<td><strong>47.27</strong></td>
<td><strong>3.64</strong></td>
<td><strong>9.09</strong></td>
<td><strong>3.64</strong></td>
<td><strong>10.91</strong></td>
<td><strong>23.64</strong></td>
<td><strong>55</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**AUTHOR’S NOTES:**
When we make the respondent aware of that they previously have failed the question and ask them to thoroughly read it again 47 percent now succeed on the question while **53 percent still fail** on it. Which means that there are quite many who actually do not read either of the questions.
1. Motivational Questions

<table>
<thead>
<tr>
<th></th>
<th>First Motivational: Percent leaving comment</th>
<th>Second Motivational: Percent leaving comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Awareness Control + Motivation</td>
<td>84.5</td>
<td>81.8</td>
</tr>
<tr>
<td>Awareness Control + Force + Motivation</td>
<td>91.9</td>
<td>87.0</td>
</tr>
<tr>
<td>Awareness Control + Force + Motivation</td>
<td>89.4</td>
<td>85.7</td>
</tr>
<tr>
<td>Total</td>
<td>88.6</td>
<td>85.0</td>
</tr>
</tbody>
</table>

AUTHOR’S NOTE:
For the motivational questions roughly 88.6 and 85 percent give an actual motivation while around 15 percent chose not to [respond].
1. Dropout after Awareness Control + Motivation

<table>
<thead>
<tr>
<th></th>
<th>Control Group</th>
<th>Awareness Control</th>
<th>Awareness Control + Force</th>
<th>No Awareness Control + Motivation</th>
<th>Awareness Control + Motivation</th>
<th>Awareness Control + Force + Motivation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dropout after Awareness Control</td>
<td>1.81</td>
<td>1.24</td>
<td>2.10</td>
<td>3.44</td>
<td>4.52</td>
<td>3.74</td>
<td>2.80</td>
</tr>
<tr>
<td>Finished</td>
<td>98.19</td>
<td>98.76</td>
<td>97.90</td>
<td>96.56</td>
<td>95.48</td>
<td>96.26</td>
<td>97.20</td>
</tr>
<tr>
<td>N</td>
<td>332</td>
<td>322</td>
<td>333</td>
<td>291</td>
<td>332</td>
<td>321</td>
<td>1,931</td>
</tr>
</tbody>
</table>
1. Time Used

Time per question:
- Decreased after Awareness control question.
- Increased after Motivational question.
### 1. Logistic Regression: Straight-lining

<table>
<thead>
<tr>
<th></th>
<th>Weak Straight-Lining</th>
<th>Strong Straight-Lining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Group</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Awareness Control</td>
<td>-0.24 (0.23)</td>
<td>-0.35 (0.31)</td>
</tr>
<tr>
<td>Awareness Control + Force</td>
<td>-0.02 (0.22)</td>
<td>-0.01 (0.28)</td>
</tr>
<tr>
<td>No Awareness Control + Motivation</td>
<td>-0.06 (0.23)</td>
<td>-0.09 (0.30)</td>
</tr>
<tr>
<td>Awareness Control + Motivation</td>
<td>-0.36 (0.24)</td>
<td>-0.62 (0.33)</td>
</tr>
<tr>
<td>Awareness Control + Force + Motivation</td>
<td>0.02 (0.22)</td>
<td>0.15 (0.28)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.75*** (0.16)</td>
<td>-2.40*** (0.20)</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>1889</td>
<td>1883</td>
</tr>
</tbody>
</table>

**Does satisficing decrease?**

**AUTHOR: “No”**
2. Removing Failers - Logistic Regression: Straight-lining

<table>
<thead>
<tr>
<th></th>
<th>With Failers Weak Straight-Lining</th>
<th>With Failers Strong Straight-Lining</th>
<th>Without Failers Weak Straight-Lining</th>
<th>Without Failers Strong Straight-Lining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness Control</td>
<td>-0.24 (0.23)</td>
<td>-0.35 (0.31)</td>
<td>-0.26 (0.24)</td>
<td>-0.40 (0.32)</td>
</tr>
<tr>
<td>Awareness Control + Force</td>
<td>-0.02 (0.22)</td>
<td>-0.01 (0.28)</td>
<td>-0.04 (0.23)</td>
<td>-0.07 (0.29)</td>
</tr>
<tr>
<td>No Awareness Control + Motivation</td>
<td>-0.06 (0.23)</td>
<td>-0.09 (0.30)</td>
<td>-0.06 (0.23)</td>
<td>-0.09 (0.30)</td>
</tr>
<tr>
<td>Awareness Control + Motivation</td>
<td>-0.36 (0.24)</td>
<td>-0.62 (0.33)</td>
<td>-0.40 (0.25)</td>
<td>-0.68 (0.35)</td>
</tr>
<tr>
<td>Awareness Control + Force + Motivation</td>
<td>0.02 (0.22)</td>
<td>0.15 (0.28)</td>
<td>-0.08 (0.23)</td>
<td>-0.00 (0.29)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.75*** (0.16)</td>
<td>-2.40*** (0.20)</td>
<td>-1.75*** (0.16)</td>
<td>-2.40*** (0.20)</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>1889</td>
<td>1883</td>
<td>1795</td>
<td>1789</td>
</tr>
</tbody>
</table>

Was data quality better for those who passed the test? 

**AUTHOR: no difference**
## 4. Low- versus High-Ability Logistic Regression: Straight-lining

<table>
<thead>
<tr>
<th></th>
<th>Weak Straight-Lining</th>
<th>Strong Straight-Lining</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low-Ability</td>
<td>Low-Ability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High-Ability</td>
</tr>
<tr>
<td>Control Group</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.</td>
</tr>
<tr>
<td>Awareness Control</td>
<td>0.03</td>
<td>-0.17</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.42)</td>
</tr>
<tr>
<td>Awareness Control + Force</td>
<td>0.51</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
<td>(0.39)</td>
</tr>
<tr>
<td>No Awareness Control + Motivation</td>
<td>0.24</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.34)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>Awareness Control + Motivation</td>
<td>-0.21</td>
<td>-0.59</td>
</tr>
<tr>
<td></td>
<td>(0.36)</td>
<td>(0.46)</td>
</tr>
<tr>
<td>Awareness Control + Force + Motivation</td>
<td>0.29</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>(0.33)</td>
<td>(0.38)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.92***</td>
<td>-2.34***</td>
</tr>
<tr>
<td></td>
<td>(0.25)</td>
<td>(0.29)</td>
</tr>
<tr>
<td>N</td>
<td>920</td>
<td>919</td>
</tr>
</tbody>
</table>

**Note:** Education level has no impact on data quality.
Conclusion

- Introducing an Awareness Control early in the survey and asking respondents to motivate their answers did not increase dropout

- Introducing an Awareness Control did not decrease satisficing

- Getting respondents to motivate their answers did not decrease satisficing

- Introducing an Awareness Control made respondents answer questions quicker and Motivation Questions slowed them down afterwards

- Additionally, if anything, High-Ability Respondents decreased in Concurrent Validity after the Awareness Control

- Controlling respondents Awareness at one time in the survey does not seem to predict overall bad survey response behavior or satisficing behavior.

- In short: Do not include Awareness Control or Motivational Questions to increase Data Quality.
• Validity testing helps identify differences resulting from method, context, interpretation, and respondent (Scanlon)

• Use of Proxies increases response but is it worth it considering the variances in answers? (Medway & Payne)

• When similar surveys yield different results, the explanation should be identified & considered before actions are taken. (Robinson & Triplett)

• Introducing “speed controls” or “pauses” in the survey does not decrease satisficing (Lundmark)