



69th Annual
Conference

Measurement
and the Role of
Public Opinion
in a Democracy

May 15 – 18, 2014
Anaheim Marriott • Anaheim, CA
www.aapor.org

#AAPOR

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



Paul Scanlon, PhD

Research Social Scientist and Survey Methodologist
Questionnaire Design Research Lab
National Center for Health Statistics
Centers for Disease Control and Prevention

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)



U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

SAFER • HEALTHIER • PEOPLE™



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?

Interpretation Matters

- 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” ...

Respondent Differences Matter

- 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

Sterilizing Medical Instruments

50.1 Which of the following sterilants or high-level disinfectants do you use to sterilize medical instruments?

Acetic acid
 No Yes Don't Know

Alcohol, such as ethanol or isopropanol
 No Yes Don't Know

Bleach or chlorine, such as Clorox®
 No Yes Don't Know

Enzymatic cleaners, such as Asepti-Zyme®, 3M Rapid Multi-Enzyme®
 No Yes Don't Know

Ethylene oxide in compressed-gas cylinders, single-dose cartridges or glass ampules
 No Yes Don't Know

Formaldehyde
 No Yes Don't Know

Glutaraldehydes such as Cidex®, Metricide®, Rapicide®, Wavicide®, Aldahol III®, Sporidicin
 No Yes Don't Know

Hydrogen peroxides such as Accell®, Optim®, Resert XL®, Sporox®, Acecide®, EndoSpore Plus®, Metrex®, Peract®, Sterad®
 No Yes Don't Know

Hypochlorite or hypochlorous acids such as Sterilox®
 No Yes Don't Know

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



Julianne Payne, PhD

Research Analyst,
Social & Scientific Systems

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

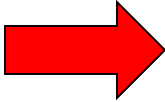
Research Questions

**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

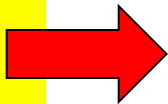
 **Target = physician; Proxy = office staff**

Results: Item Missingness

Domain	% Missing, Target	% Missing, Proxy
Survey Focus	23.9	7.4
Org Structure	11.5	4.3
Administration	23.2	6.9
Clinical Treatment	23.1	6.9

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

Proxies answered more completely



**13.2% missing among physicians
3.1% missing among proxies**

Results: Content of Response, Study Focus

On Focus: Proxies and Targets gave different answers

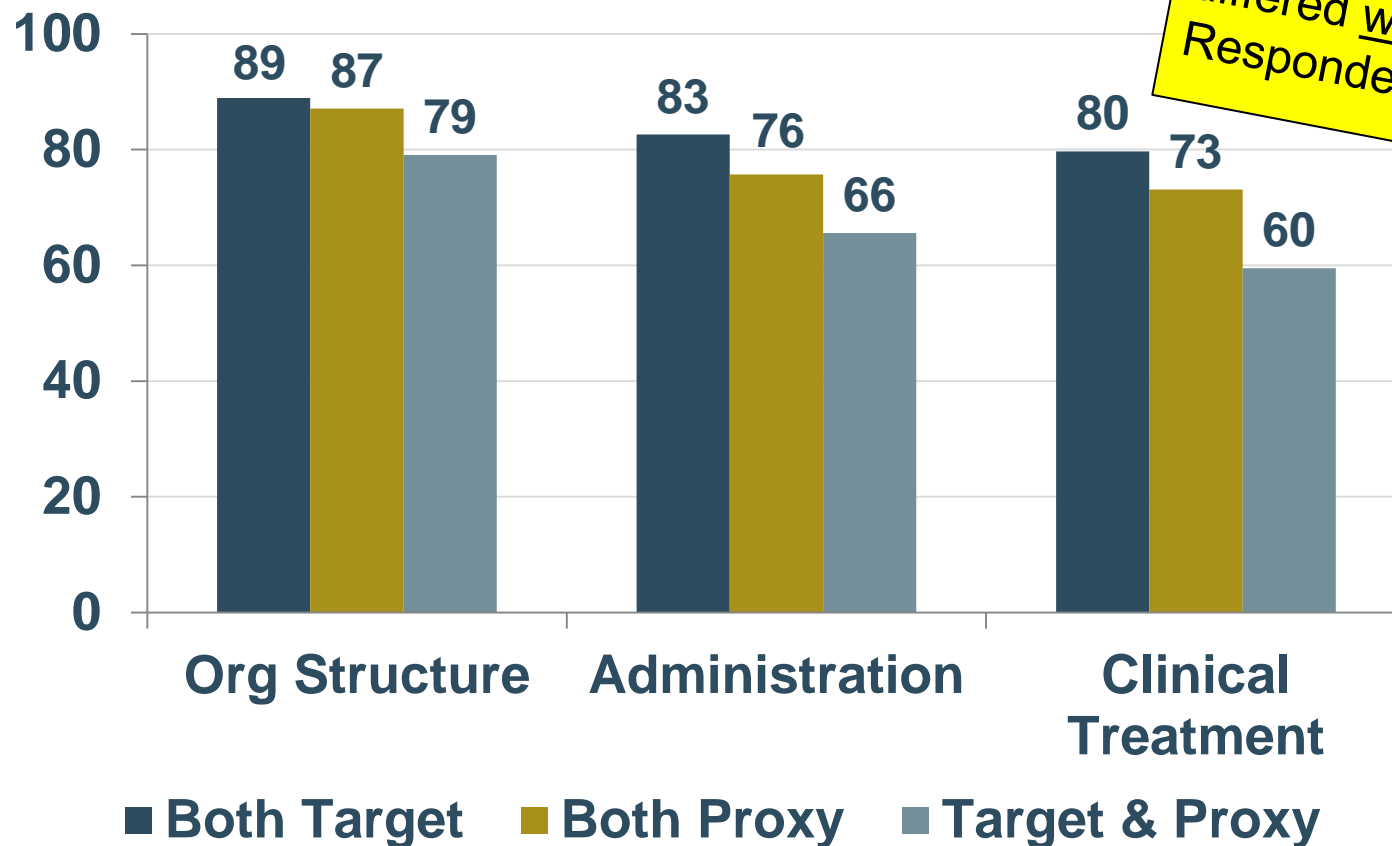
Survey Question	% of Matching Responses, Missing as Valid			% of Matching Responses, Missing as Invalid		
	Both Target	Both Proxy	Target & Proxy	Both Target	Both Proxy	Target & Proxy
Practice uses EHR	84.8	83.7	68.0	87.5	88.4	79.6
Name of EHR	89.2	85.7	72.0	96.0	100.0	94.6
Year EHR installed	75.2	79.6	60.8	70.0	73.9	58.6
Plans to install EHR	79.0	67.3	56.0	89.1	85.3	69.1

 **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



.... 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

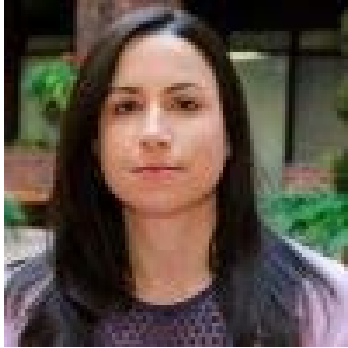
- **High within-type variation raises questions**

- 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**



Rebecca Medway, PhD

Survey Methodologist
Center for Survey Methods,
American Institutes for Research

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

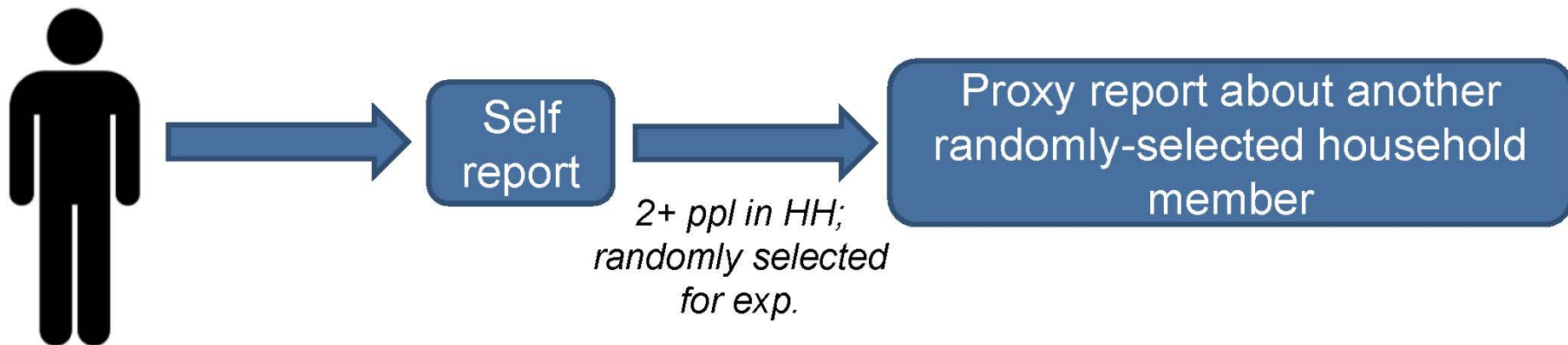
Rebecca Medway and Celeste Stone

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)

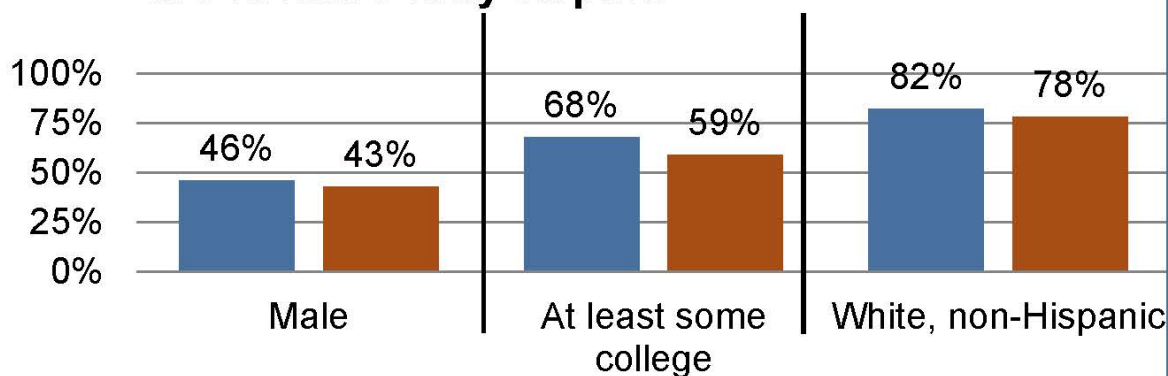


Results:

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



- Completed proxy report (n=1040)
- Did not end up completing proxy report (n=120)

Reporter Mean Age by Willingness to Provide Proxy Report

Interview type	Mean age
Completed proxy report (n=1040)	50.7
Did not complete proxy report (n=120)	54.2

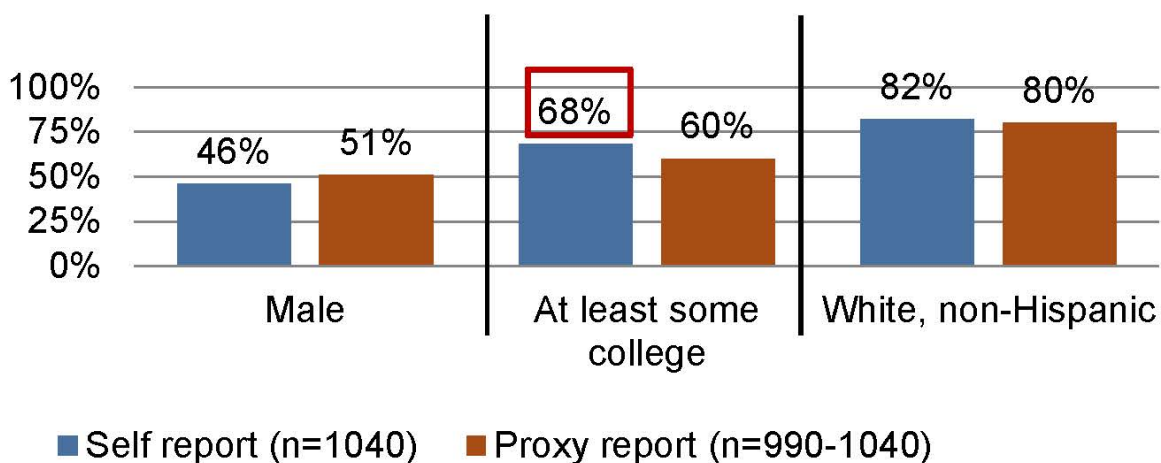
 Significant difference ($p < 0.05$)

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.

Subject Demographics by Report Type



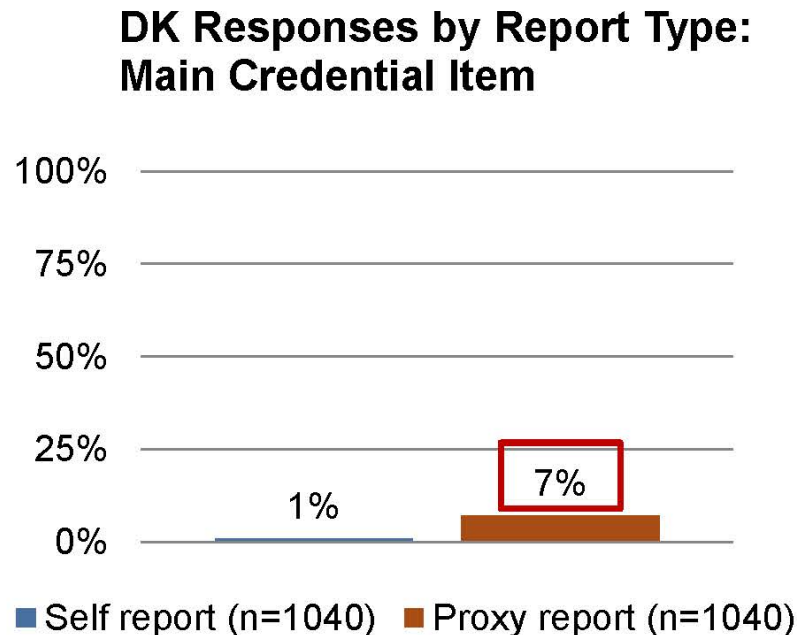
Reporter Mean Age by Report Type

Interview type	Mean age
Self report (n=1040)	50.7
Proxy report (n=1040)	49.9

□ Significant difference btwn self and proxy ($p < 0.05$)

Results: How Well – DK Response to Main Credential Item

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

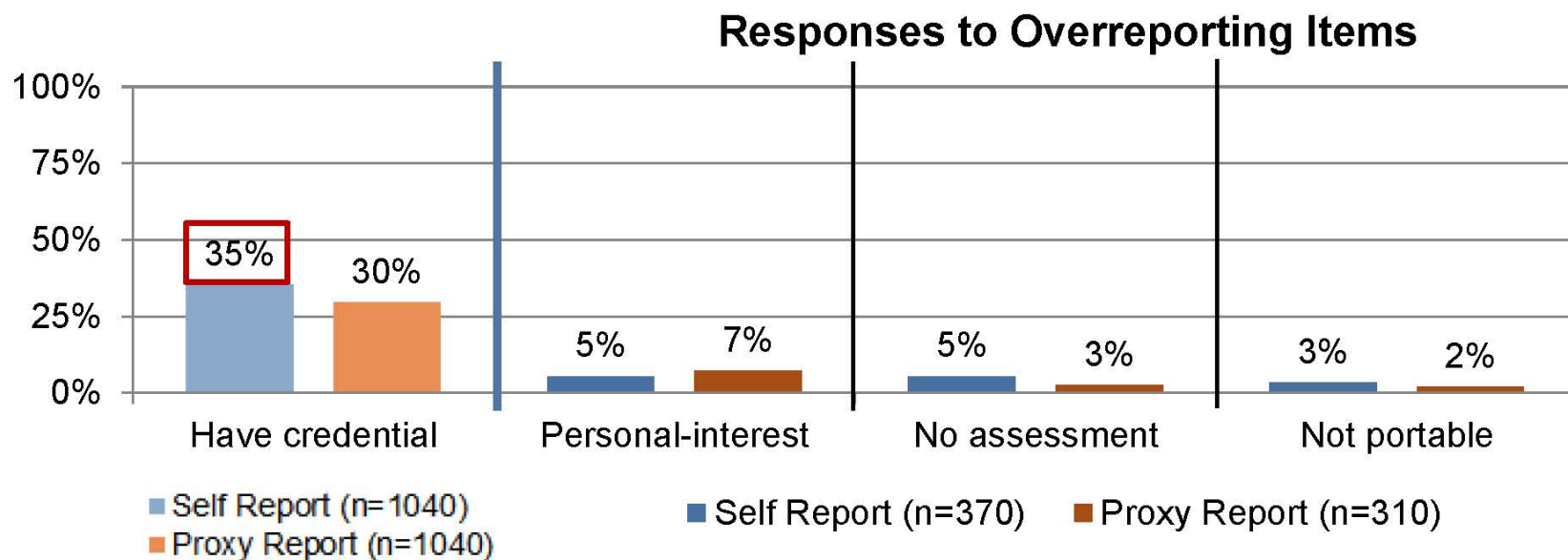


True regardless of age or education

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).



Significant difference btwn 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?



John Robinson, PhD

Department of Sociology
University of Maryland



Timothy Triplett

Urban Institute
2100 M Street, NW
Washington, DC 20037

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.



69th Annual
Conference

Measurement
and the Role of
Public Opinion
in a Democracy

May 15 – 18, 2014

Anaheim Marriott • Anaheim, CA

www.aapor.org

#AAPOR

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

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

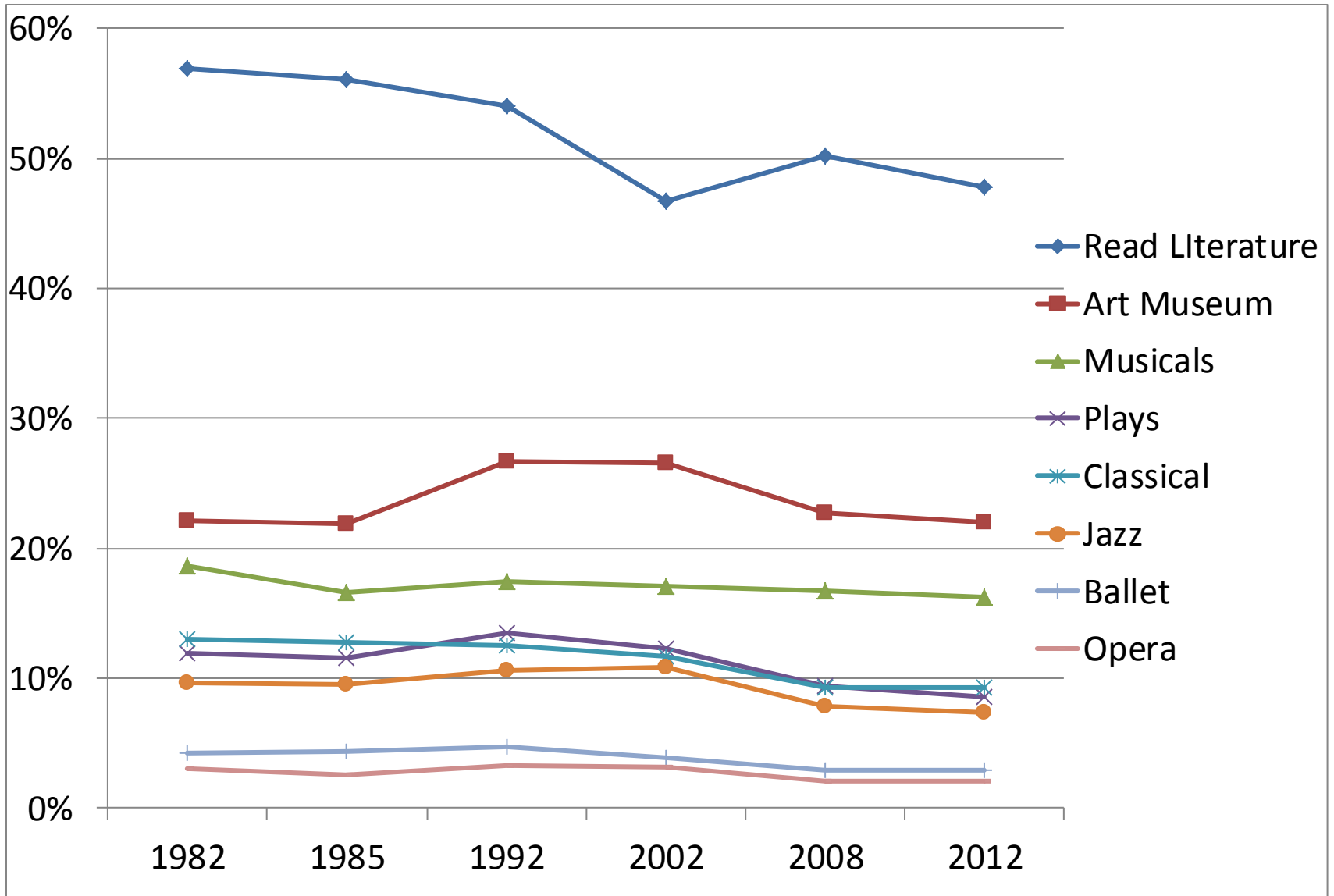
Arts Participation Questions

(Source: Census)

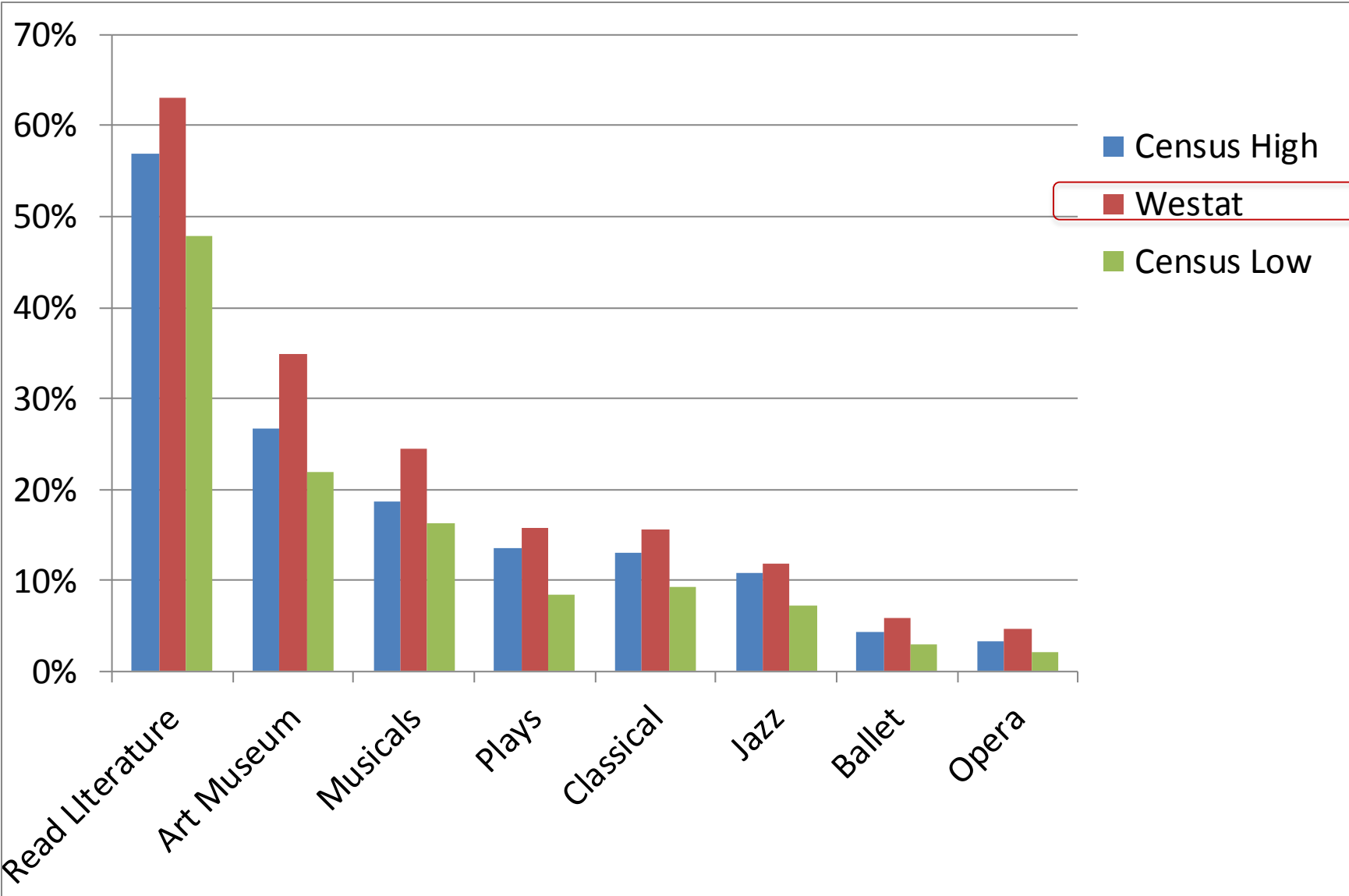
- With the exception of elementary or high school performances, did you go to a live [fill below] performance between **{today's date}** 2011 and **{today's date}** 2012?
 - Jazz
 - Classical music such as symphony, chamber or choral music
 - Opera
 - Musical stage play
 - Non-Musical stage play
 - Ballet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



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.



Sebastian Lundmark

Doctoral Candidate

Department of Political Science
University of Gothenburg, Sweden

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.”*

Strongly
Disagree

1



2



3



4



5



6



Strongly
Agree

7





After the Awareness Control question, open ends are added to attempt to increase the respondent's focus.

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

	Strongly Disagree 1	(Correct) 2	3	4	5	6	Strongly Agree 7	n
Awareness Control	0.00	93.23	0.00	0.97	0.97	1.61	3.23	310
Awareness Control + Force	0.00	93.01	0.91	1.22	0.91	0.91	3.04	329
Awareness Control + Motivation	0.62	91.36	0.31	1.54	0.62	2.47	3.09	324
Awareness Control + Force + Motivation	0.32	91.77	0.00	2.53	0.63	2.53	2.22	316
Total	0.23	92.34	0.31	1.56	0.78	1.88	2.89	1279

1. Results – Failed/Successful of Awareness Control

Opting Out

Acquiescence

	Strongly Disagree 1	(Correct) 2	3	4	5	6	Strongly Agree 7	n
Awareness Control	0.00	93.23	0.00	0.97	0.97	1.61	3.23	310
Awareness Control + Force	0.00	93.01	0.91	1.22	0.91	0.91	3.04	329
Awareness Control + Motivation	0.62	91.36	0.31	1.54	0.62	2.47	3.09	324
Awareness Control + Force + Motivation	0.32	91.77	0.00	2.53	0.63	2.53	2.22	316
Total	0.23	92.34	0.31	1.56	0.78	1.88	2.89	1279

1. Awareness Control + Force Follow-up Question

	Strongly Disagree 1	(Correct) 2	3	4	5	6	Strongly Agree 7	n
Awareness Control + Force	0.00	40.00	8.00	4.00	4.00	12.00	32.00	25
Awareness Control + Force + Motivation	3.33	53.33	0.00	13.33	3.33	10.00	16.67	30
Total	1.82	47.27	3.64	9.09	3.64	10.91	23.64	55

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

	First Motivational: Percent leaving comment	Second Motivational: Percent leaving comment
No Awareness Control + Motivation	84.5	81.8
Awareness Control + Force + Motivation	91.9	87.0
Awareness Control + Force + Motivation	89.4	85.7
Total	88.6	85.0

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

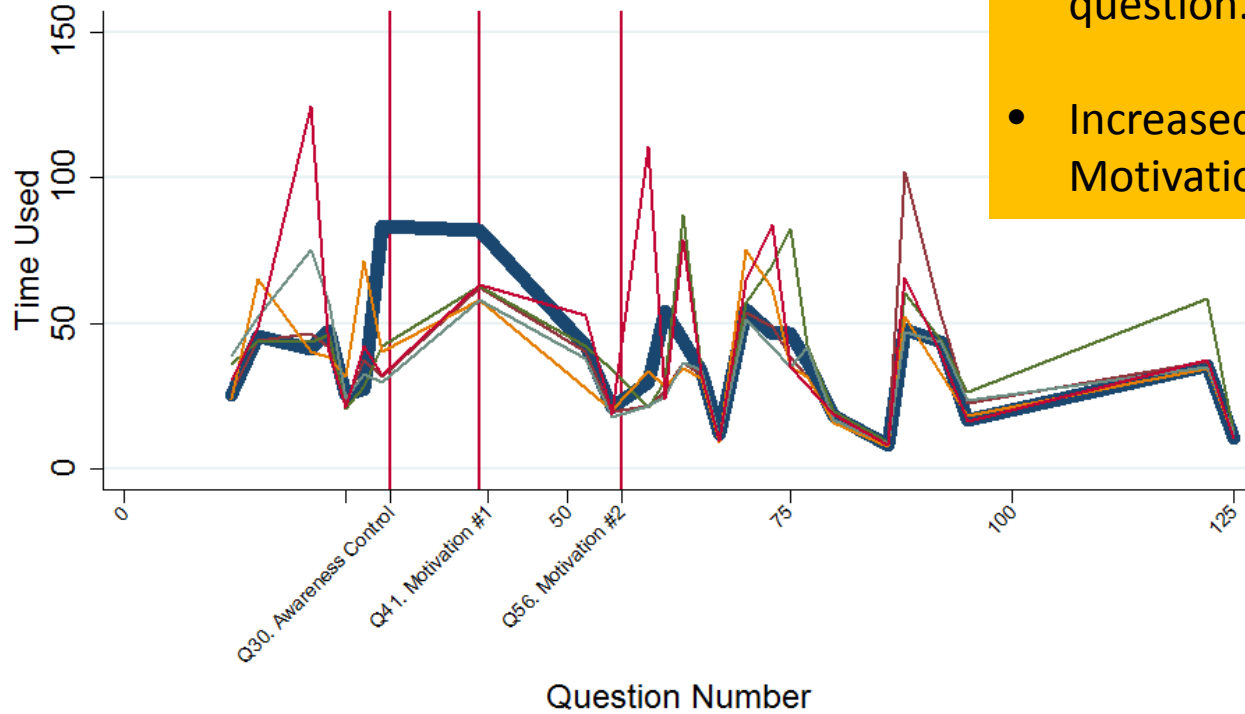
	Control Group	Awareness Control	Awareness Control + Force	No Awareness Control + Motivation	Awareness Control + Motivation	Awareness Control + Force + Motivation	Total
Dropout after Awareness Control	1.81	1.24	2.10	3.44	4.52	3.74	2.80
Finished	98.19	98.76	97.90	96.56	95.48	96.26	97.20
N	332	322	333	291	332	321	1,931

Drop out rates don't increase when Awareness Control & Motivation questions are added.

Time per question

- Decreased after Awareness control question.
- Increased after Motivational question.

1. Time Used



1. Logistic Regression: Straight-lining

	(1) Weak Straight- Lining	(2) Strong Straight- Lining
Control Group	.	.
Awareness Control	-0.24 (0.23)	-0.35 (0.31)
Awareness Control + Force	-0.02 (0.22)	-0.01 (0.28)
No Awareness Control + Motivation	-0.06 (0.23)	-0.09 (0.30)
Awareness Control + Motivation	-0.36 (0.24)	-0.62 (0.33)
Awareness Control + Force + Motivation	0.02 (0.22)	0.15 (0.28)
Constant	-1.75*** (0.16)	-2.40*** (0.20)
<i>N</i>	1889	1883

Does satisficing decrease?
AUTHOR: "No"



2. Removing Failers - Logistic Regression: Straight-lining

	With Failers Weak Straight-	With Failers Strong Straight-	Without Failers Weak Straight-	Without Failers Strong Straight-
<p>Was data quality better for those who passed the test? AUTHOR: no difference</p>				
Control Group				
Awareness Control	-0.24 (0.23)	-0.35 (0.31)	-0.26 (0.24)	-0.40 (0.32)
Awareness Control + Force	-0.02 (0.22)	-0.01 (0.28)	-0.04 (0.23)	-0.07 (0.29)
No Awareness Control + Motivation	-0.06 (0.23)	-0.09 (0.30)	-0.06 (0.23)	-0.09 (0.30)
Awareness Control + Motivation	-0.36 (0.24)	-0.62 (0.33)	-0.40 (0.25)	-0.68 (0.35)
Awareness Control + Force + Motivation	0.02 (0.22)	0.15 (0.28)	-0.08 (0.23)	-0.00 (0.29)
Constant	-1.75*** (0.16)	-2.40*** (0.20)	-1.75*** (0.16)	-2.40*** (0.20)
N	1889	1883	1795	1789



4. Low- versus High-Ability

- Logistic Regression: Straight-lining

Education level has no impact on data quality

	(1) Weak Straight-Lining Low-Ability	(2) Strong Straight-Lining Low-Ability	(3) Weak Straight-Lining High-Ability	(4) Strong Straight-Lining High-Ability
Control Group
Awareness Control	0.03 (0.34)	-0.17 (0.42)	-0.49 (0.33)	-0.62 (0.48)
Awareness Control + Force	0.51 (0.32)	0.23 (0.39)	-0.59 (0.33)	-0.34 (0.43)
No Awareness Control + Motivation	0.24 (0.34)	0.15 (0.40)	-0.35 (0.33)	-0.47 (0.48)
Awareness Control + Motivation	-0.21 (0.36)	-0.59 (0.46)	-0.47 (0.32)	-0.66 (0.48)
Awareness Control + Force + Motivation	0.29 (0.33)	0.41 (0.38)	-0.23 (0.31)	-0.23 (0.43)
Constant	-1.92*** (0.25)	-2.34*** (0.29)	-1.62*** (0.20)	-2.45*** (0.28)
N	920	919	964	959



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.**



69th Annual
Conference

Measurement
and the Role of
Public Opinion
in a Democracy

May 15 – 18, 2014

Anaheim Marriott • Anaheim, CA

www.aapor.org

#AAPOR

- **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)