

# IMPACT OF RESPONSE STYLES ON INCLUSIVE MEASUREMENT

FERNANDA ALVARADO-LEITON, PH.D., UNIVERSITY OF COSTA RICA

RACHEL DAVIS, PH.D., UNIVERSITY OF SOUTH CAROLINA

SUNGHEE LEE, PH.D., UNIVERSITY OF MICHIGAN

INCLUSIVE RESEARCH MATTERS SEMINAR SERIES

UNIVERSITY OF MICHIGAN

OCTOBER 27, 2021

# The Problem

- Many social science constructs – e.g., attitudes and beliefs – are unobservable and, thus, difficult to measure
- These types of constructs are often assessed in surveys using response scales in which respondents rate statements along a continuum, such as this disagree/agree scale:
  - Ex: I am confident that I can eat vegetables when there are foods in my house like chips, cookies, or candy. Please choose a number between 1 and 7, where 1 means “strongly disagree” and 7 means “strongly agree.”
- However, past research suggests that **different social groups tend to engage in different patterns of responding** to these types of rating scales ...

# Response Styles

- **Response styles** = Systematic patterns of responding to survey questions, regardless of item content
  - **Extreme response style (ERS)** = A systematic tendency to select response scale endpoints ●○○○○●
  - **Middle response style (MRS)** = A systematic tendency to select middle responses ○○○●○○○
  - **Disacquiescent response style (DARS)** = A systematic tendency to select responses indicating disagreement ●●●○○○○
  - **Acquiescent response style (ARS)** = Systematic agreement with survey items, regardless of item content or directionality ○○○○●●●
  - And many more ...

# Response Styles

- **Response styles** = Systematic patterns of responding to survey questions, regardless of item content
  - **Extreme response style (ERS)** = A systematic tendency to select response scale endpoints ●○○○○●
  - **Middle response style (MRS)** = A systematic tendency to select middle responses ○○○●○○○
  - **Disacquiescent response style (DARS)** = A systematic tendency to select responses indicating disagreement ●●●○○○
  - **Acquiescent response style (ARS)** = Systematic agreement with survey items, regardless of item content or directionality ○○○○●●●
  - And many more ...

# Why Is ARS a Problem?



# Ex: Measuring Life Satisfaction – 1

## 1 item measure

Please think about your life-as-a-whole. How satisfied are you with it?

### Response options:

1. Not at all satisfied
2. Not very satisfied
3. Somewhat satisfied
4. Very satisfied
5. Completely satisfied

## 5 item measure

Please say how much you agree or disagree with the following statements.

Q1. In most ways my life is close to ideal.

Q2. The conditions of my life are excellent.

Q3. I am satisfied with my life.

Q4. So far, I have gotten the important things I want in life.

Q5. If I could live my life again, I would change almost nothing.

### Response options:

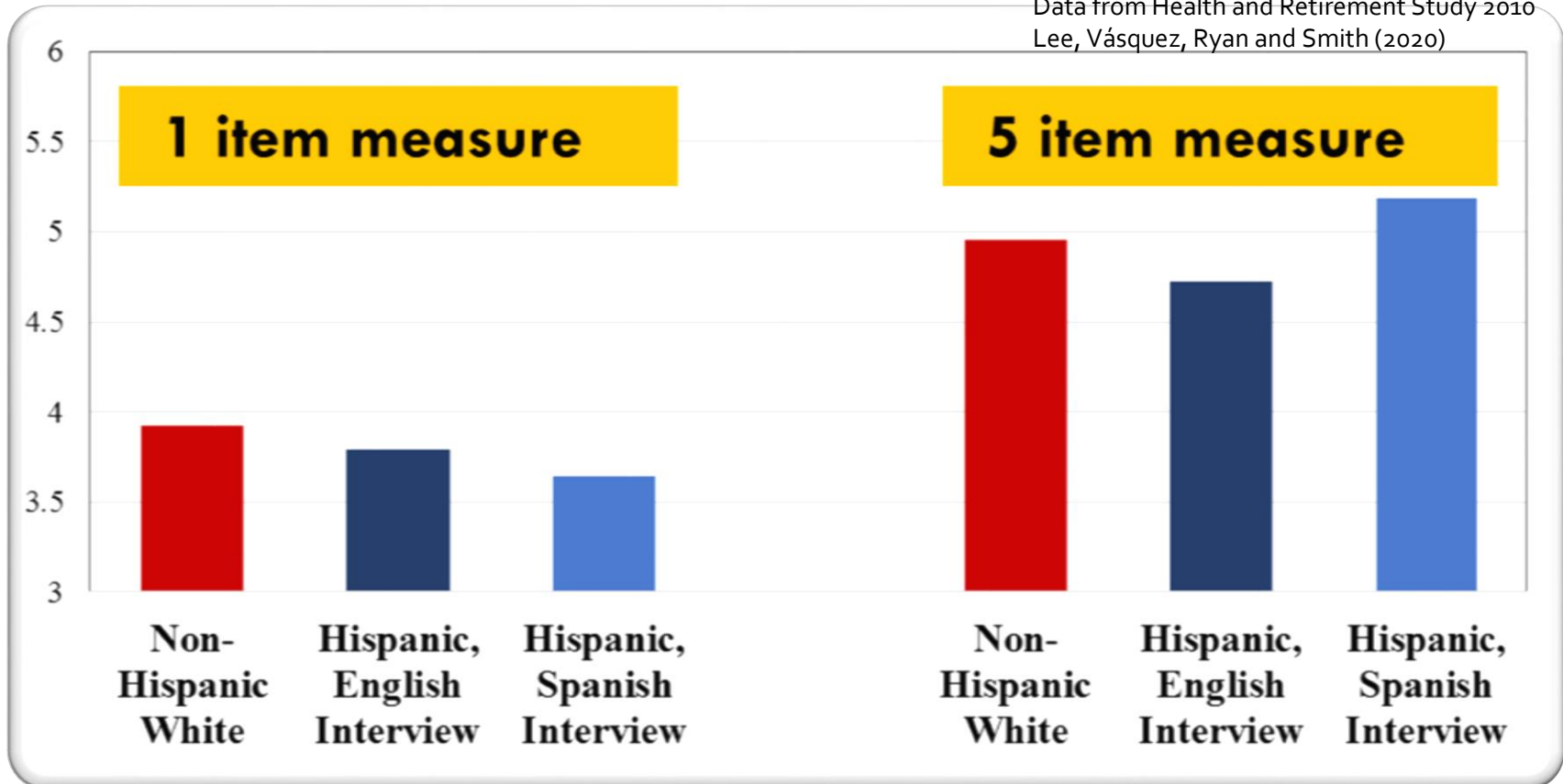
1. Strongly disagree
2. Somewhat disagree
- .....
6. Somewhat agree
7. Strongly agree



**Likert  
Scale!**

# Ex: Measuring Life Satisfaction – 2

Data from Health and Retirement Study 2010  
Lee, Vásquez, Ryan and Smith (2020)



**Acquiescent response style -- Group specific**

# ARS: A Threat to Data Quality

- ARS can lead to error in computing scores across cultural groups, reliability statistics, relationships between scales, model fit statistics, differences between ARS-adjusted and unadjusted scores, and factor structures
- **ARS threatens scientific knowledge** – BUT it does so in inequitable ways across social groups
- While ARS may be a small and nonsignificant source of error in some surveys, it may be a substantial source of error when surveying populations with a tendency to acquiesce
- Thus, ARS is more impactful in research with acquiescent populations, who also tend to be populations of particular interest for many social scientists

E.g., Billiet & Davidov, 2008; Billiet & McClendon, 2000; Cheung & Rensvold, 2000; Lee, Vasquez, Ryan, & Smith, 2020; Weijters, Schillewaert, & Geuens, 2008



# Social Patterning of ARS

- Among other factors, previous studies suggest that ARS tends to be higher among:
  - Respondents with less education
  - Older respondents
  - Latino and African American respondents in the U.S., as compared to non-Latino white respondents
  - Respondents from collectivist cultures

# References

- Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7(4), 608-628.
- Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research*, 36(4), 542-562.
- Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2), 187-212.
- Davis, R. E., Johnson, T. P., Lee, S., & Werner, C. (2019). Why do Latino survey respondents acquiesce? Respondent and interviewer characteristics as determinants of cultural patterns of acquiescence among Latino survey respondents. *Cross-Cultural Research*, 53(1), 87-115.
- Lee, S., Vasquez, E., Ryan, L., & Smith, J. (2020). Measurement equivalence of subjective well-being scales under the presence of acquiescent response style for the racially and ethnically diverse older population in the United States. *Survey Research Methods*, 14(4), 417-437.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2), 195-217.
- Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science*, 36(3), 409-422.

# REDUCING ACQUIESCENT RESPONSE STYLE WITH CONVERSATIONAL INTERVIEWING

RACHEL DAVIS, FREDERICK CONRAD, SHAOHUA DONG, ANNA MESA,  
SUNGHEE LEE, & TIMOTHY JOHNSON

INCLUSIVE RESEARCH MATTERS SEMINAR SERIES  
UNIVERSITY OF MICHIGAN  
OCTOBER 27, 2021

# ARS among U.S. Latinos

- The Latino population comprises approximately 19% of the U.S. population
- Studies indicate that Latino adults tend to acquiesce more than non-Latino whites
- Unclear if this is due to differences in sociodemographic characteristics, cultural norms guiding communication and/or social interactions, or something else
- **ARS can lead to erroneous conclusions about Latino populations** and artificially enhance or decrease distinctions between Latino populations and other social groups

# Addressing ARS

- It is sometimes possible to control for ARS-induced error after data collection during the data analysis phase
- However, **it would be preferable to reduce the occurrence of ARS at the time of data collection**, as this should produce more valid raw responses and reduce the need for later statistical adjustments



# Research Questions

- Using a 2x2 experimental design, we administered a telephone survey with acquiescent Latino adults to address the following research questions:
  1. Will using **conversational interviewing** methods to administer a survey questionnaire yield lower ARS than a standardized interviewing approach?
  2. Will using unipolar, **item-specific response scales** yield lower ARS than using bipolar, disagree/agree scales?

# Standardized Interviewing

- **Standardized interviewing (SI)** = Interviewers are required to read survey questions exactly as worded; if a respondent provides something other than an acceptable answer, interviewers can use “neutral” probes to obtain an acceptable response
- The goal of SI is to **hold question wording constant** across respondents so that different answers can be attributed to differences between respondents rather than to differences in question wording

# Conversational Interviewing

- **Conversational interviewing (CI)** = Interviewers are trained to provide assistance to respondents; interviewers read questions exactly as worded, judge whether respondents need clarification, and say what they believe is necessary to convey the intended meaning of the question
- The wording that interviewers use is allowed to vary across respondents
- The goal of CI is to **standardize respondents' interpretation of each question** rather than the question wording itself



# Ex: Standardized Interviewing

- **Interviewer:** *In some situations, it is more important to be compassionate than fair.*
- **Respondent:** *What do you mean by "compassionate"?*
- **Interviewer:** *Whatever it means to you.*

# Ex: Conversational Interviewing

- **Interviewer:** *In some situations, it is more important to be compassionate than fair.*
- **Respondent:** *What do you mean by "compassionate"?*
- **Interviewer:** *This question is asking whether you believe that a person should always be fair to other people, no matter what, or whether there are sometimes situations in which it is more important to be caring, thoughtful, considerate, and sympathetic toward other people.*

# How Might CI Reduce ARS?

- CI interviewers can **provide personalized assistance** to address the specific difficulties experienced by each respondent
  - Ex: While one respondent may have difficulty understanding the wording used in a particular survey item, another respondent may have difficulty with response mapping
- CI interviewers can **convey the importance of providing honest, thoughtful answers**
  - Ex: In a previous study, we asked Mexican American participants why they thought Mexican Americans might be more likely to engage in ARS. One participant commented:  
*(W)e have been programmed that, "Well, I guess they just wanna hear the positive, and they don't want to hear what we really have to say" ... if you ain't got nothing good to say, don't say it ...*

# Disagree/Agree Response Scales

- **Disagree/agree (DA) response scales** = Include labels ranging from disagreement (e.g., “Strongly disagree”) to agreement (e.g., “Strongly agree”):
  - Ex: *I am confident that I can prepare a meal that contains vegetables when I am tired. Please choose a number between 1 and 7, where 1 means “strongly disagree” and 7 means “strongly agree.”*
- Tend to be bipolar (e.g., disagreement to agreement or vice versa)
- Ranging from disagree to agree may be associated with higher ARS

# Item-Specific Response Scales

- **Item-specific response scales** = Include response option labels that are intended to match the underlying construct assessed in each survey item:
  - Ex: *How **confident** are you that you can prepare a meal that contains vegetables when you are tired? Please choose a number between 1 and 7, where 1 means "not at all **confident**" and 7 means "very confident."*
- Are often unipolar (e.g., low to high confidence)

# How Might Item-Specific Response Scales Reduce ARS?

- Item-specific scales **do not explicitly ask about agreement**
  - DA scales explicitly ask about agreement, which may motivate respondents to select agreement responses to convey deference, express politeness, or avoid embarrassment due to difficulty responding to the question
- Item-specific scales may be associated with **decreased cognitive burden**
  - DA scales require respondents to engage in an extra mental step of mapping their answers onto an agreement dimension, which, in most cases, is not the dimension underlying the item (Ex: the previous example that assessed “confidence”)
- Item-specific scales may be **less obvious about which end of the scale is positive or negative**
  - DA scales (Ex: “Strongly Disagree”/“Strongly Agree”) require respondents to select a response from the negative or positive end of the scale
  - Since item-specific scales are often unipolar, their values are not as clearly labeled (Ex: “Believe this very much” is not necessarily more positive than “Don’t believe this at all”)

# Participants

- **891 telephone survey respondents:**
  - Eligibility criteria: Aged 18-90; spoke English or Spanish; Mexican American, Cuban American, or Puerto Rican heritage, and exhibited a threshold level of ARS
  - Interviews conducted in Spanish and English
  - Threshold ARS: Answered 6 or 7 on a scale ranging from 1="Strongly Disagree" to 7="Strongly Agree" to  $\geq 1$  items in both the positively and negatively worded item groups comprising two, 10-item balanced scales
  - Recruited using a commercial sample vendor list of landline and mobile phone numbers associated with addresses in large Latino markets in the mainland U.S. or Puerto Rico, lower-income households (\$25,000 or less), and individuals with limited education (12 years or less)

# Interviewers

- **28 professional interviewers:**
  - 22 SI interviewers; 6 CI interviewers
  - Located in different call centers and unaware of the other protocol
- Both sets of interviewers were trained to administer the questionnaires using their assigned interviewing technique
- **CI training:**
  - Focused on teaching the intended meanings of the questions
  - Provided interviewers with a binder containing definitions for all 132 questions
  - Taught how the response scales were intended to be used
  - Instructed interviewers to first read each question as worded and then say whatever they considered necessary, if anything, to ensure respondents understood the intended question meanings and how to use the response scales



# Experimental Design

- A **2x2 between-subjects design** that fully crossed interviewing technique (CI vs. SI) with response scale format (unipolar item-specific vs. bipolar DA)
- Participants were **randomly assigned to 4 groups**:
  1. SI/item-specific response options (n=301)
  2. SI/DA response options (n=295)
  3. CI/item-specific response options (n=149)
  4. CI/DA response options (n=146)
- More respondents assigned to SI (n=596) than CI (n=295) because the CI required more interviewer training, more monitoring, longer interviews, and more money
- Used **audio recordings to code interviewer behaviors** for 616 interviews (n=223 CI; n=393 SI) for each administration of the 27 items used to measure ARS

# Questionnaire

- **132 items** querying vegetable consumption, attitudes and beliefs related to vegetable consumption, and sociodemographics
- **Calculated ARS variable:**
  - 27 items regarding beliefs about unrelated topics selected to represent a heterogeneous group of items for measuring ARS (e.g., “The United States spends too much money on scientific research”)
  - 7-point response scales:
    - DA: 1=“Strongly Disagree” to 7=“Strongly Agree”
    - Item-specific: 1=“Don’t Believe This At All” to 7=“Believe This Very Much”
  - Recoded negatively valenced responses (1, 2, 3, 4) to 0 and positively valenced responses (5, 6, 7) to 1, 2, and 3, respectively; averaged the recoded values across the 27 items for each respondent
  - Values ranged from 0 to 3, with larger values indicating greater ARS

# Sample Characteristics

|                                                    | <b>Total<br/>(n=891)</b> | <b>CI Only<br/>(n=295)</b> | <b>SI Only<br/>(n=596)</b> |
|----------------------------------------------------|--------------------------|----------------------------|----------------------------|
| <b>Mean age in years (standard deviation [SD])</b> | 64.3 (15.5)              | 65.7 (15.1)                | 63.7 (15.7)                |
| <b>Gender (% female)</b>                           | 80.1                     | 81.0                       | 79.7                       |
| <b>Latinx heritage group (%):</b>                  |                          |                            |                            |
| Mexican American                                   | 34.6                     | 34.2                       | 34.7                       |
| Puerto Rican                                       | 32.9                     | 33.2                       | 32.7                       |
| Cuban American                                     | 32.6                     | 32.5                       | 32.6                       |
| <b>Education (%):</b>                              |                          |                            |                            |
| Less than 7th grade                                | 26.8                     | 28.5                       | 26.0                       |
| 7th through 12th grade, no diploma                 | 24.9                     | 23.7                       | 25.5                       |
| High school graduate or equivalent                 | 23.0                     | 20.7                       | 24.2                       |
| Some college or technical/vocational school        | 12.4                     | 11.9                       | 12.6                       |
| 4-year college degree                              | 9.1                      | 10.5                       | 8.4                        |
| Graduate degree                                    | 3.8                      | 4.8                        | 3.4                        |
| <b>Interview language (%):</b>                     |                          |                            |                            |
| Spanish                                            | 93.1                     | 94.2                       | 92.6                       |
| English                                            | 6.9                      | 5.8                        | 7.4                        |

# Fidelity to CI vs. SI Methods

- Examining the mean frequencies of the interviewer behavior codes from the audio recordings indicated that **interviewers adhered to their assigned interviewing techniques** when administering the 27 questions used to measure ARS
- On average, **interviewers assigned to SI:**
  - Provided full neutral probes in 4% of question administrations
  - Provided partial neutral probes in 16% of question administrations
  - Used non-neutral probes or definitions in <1% of question administrations
- On average, **interviewers assigned to CI:**
  - Clarified question meanings in response to evidence of respondent confusion in 12% of question administrations
  - Clarified question meanings without explicit evidence of respondent confusion in 34% of question administrations
  - Rarely (<1%) failed to provide a definition when there was evidence it may have helped
  - Helped respondents with response mapping in 29% of question administrations
  - Rarely (<1%) failed to help respondents with response mapping when assistance was needed
  - Corrected respondents' understanding of the response scale 3% of question administrations
  - Also administered neutral probes for an average of 28% of the question administrations; this use of SI methods did not appear to detract from the effectiveness of CI

# Main Findings

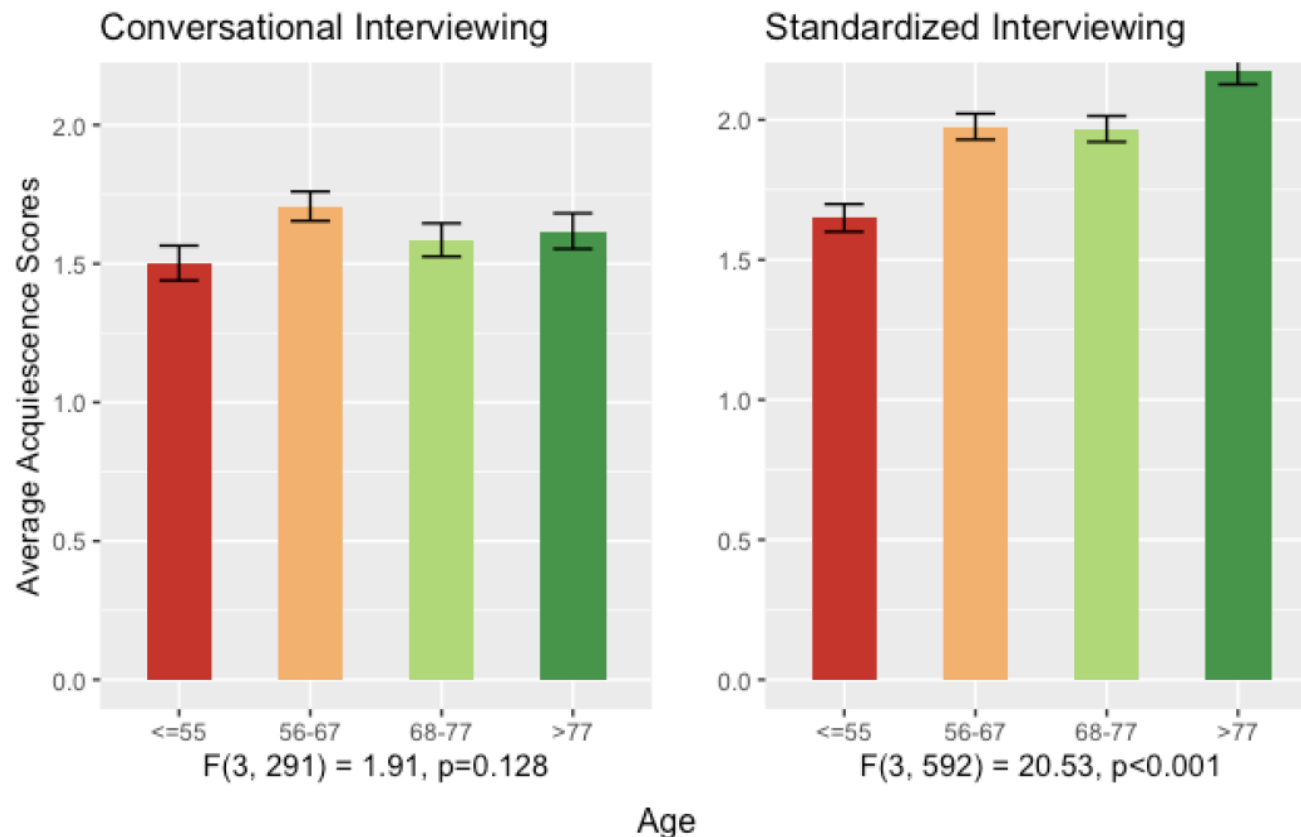
**Table 4. Linear Regression Estimates for the Influence of Interviewing Techniques (CI vs. SI) and Response Scale Formats (DA vs. Item Specific) on ARS (n=891)**

|                                                            | Model 1          |                 | Model 2          |                 |
|------------------------------------------------------------|------------------|-----------------|------------------|-----------------|
|                                                            | Estimate (SE)    | <i>p</i> -value | Estimate (SE)    | <i>p</i> -value |
| Interviewing technique (ref=SI)                            | -0.34 (0.04) *** | <0.001          | -0.34 (0.06) *** | <0.001          |
| Response scale format (ref=DA)                             | 0.04 (0.04)      | 0.23            | 0.05 (0.05)      | 0.31            |
| Latinx heritage group (ref=Mexican American):              |                  |                 |                  |                 |
| Cuban American                                             | 0.15 (0.05) #    | 0.06            | 0.15 (0.05) #    | 0.06            |
| Puerto Rican                                               | -0.09 (0.05) *** | <0.001          | -0.09 (0.05) *** | <0.001          |
| Age quartiles (ref=Younger than 55):                       |                  |                 |                  |                 |
| 55-67                                                      | 0.24 (0.05) ***  | <0.001          | 0.24 (0.05) ***  | <0.001          |
| 67-77                                                      | 0.19 (0.05) ***  | <0.001          | 0.19 (0.05) ***  | <0.001          |
| Older than 77                                              | 0.25 (0.06) ***  | <0.001          | 0.25 (0.06) ***  | <0.001          |
| Education (ref=Less than 7 <sup>th</sup> grade):           |                  |                 |                  |                 |
| 7 <sup>th</sup> through 12 <sup>th</sup> grade, no diploma | -0.10 (0.05) #   | 0.06            | -0.10 (0.05) #   | 0.06            |
| High school graduate or equivalent                         | -0.30 (0.05) *** | <0.001          | -0.30 (0.05) *** | <0.001          |
| Some college or technical/vocational school                | -0.30 (0.07) *** | <0.001          | -0.30 (0.07) *** | <0.001          |
| 4-year college degree or higher                            | -0.41 (0.07) *** | <0.001          | -0.41 (0.07) *** | <0.001          |
| Latinx cultural orientation                                | 0.11 (0.05) *    | 0.05            | 0.11 (0.05) *    | 0.05            |
| Interview language (ref=English)                           | 0.19 (0.08) *    | 0.02            | 0.19 (0.08) *    | 0.02            |
| Interviewing technique x response scale format             |                  |                 | -0.004 (0.08)    | 0.96            |
| R squared                                                  | 0.22             |                 | 0.22             |                 |
| Model <i>p</i> -value                                      | <0.001           |                 | <0.001           |                 |

# =  $p \leq 0.10$ ; \* =  $p \leq 0.05$ ; \*\* =  $p \leq 0.01$ ; \*\*\* =  $p \leq 0.001$

# Interviewing Method and Age

**Figure 1.** Analysis of Variance Comparing the Relationship between Age and ARS by Interviewing Technique (n=891)



# Effective CI Behaviors

- Analyses of the interviewer behavior codes indicate that the conversational interviewers primarily reduced ARS by:
  - Providing definitions without explicit evidence of respondent confusion
  - Helping respondents with response mapping
- Conversely, ARS *increased* when the conversational interviewers failed to help with response mapping when there was evidence that such assistance was needed
- None of the SI behaviors reduced ARS
- However, ARS *increased* when the standardized interviewers issued neutral probes in response to evidence that response mapping assistance was needed, i.e., unhelpful “help”

# Summary: DA vs. Item-Specific Scales

- We found no evidence that unipolar item-specific response scales yielded less ARS than bipolar DA scales
- It is possible that item-specific response scales have no advantage for reducing ARS over DA scales
- It is also possible that our item-specific response scales may have been insufficiently specific:
  - Ex: The response dimension “Believe this very much” was not inherently related to the topics of the questions (e.g., “global warming is a myth”)
- However, we conducted analyses with a secondary ARS measure that addressed this weakness by including items with response scales that were more closely tied to the constructs being assessed, and the same pattern of results emerged
- A second possibility is that our item-specific response scales, like traditional DA scales, were still oriented on a negative-positive continuum, making it natural for respondents to conceptualize the response options as inherently positive or negative



# Summary: CI vs. SI

- The use of CI to administer a telephone survey was associated with significantly less ARS than using SI
- CI weakened the relationship between age and ARS
- The reduction of ARS in the CI condition appeared to be primarily attributable to interviewers' efforts to clarify terms, even when respondents provided no evidence of difficulty, and provide help with response mapping
- In sum, conversational interviewing may decrease ARS during data collection, thereby reducing the need for post-survey adjustments

# References

- Conrad, F. G., & Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly*, 64(1), 1-28.
- Conrad, F. G., & Schober, M. F. (2020). Clarifying question meaning in standardized interviews can improve data quality even though wording may change: A review of the evidence. *International Journal of Social Research Methodology*, Online First.
- Davis, R. E., Johnson, T. P., Lee, S., & Werner, C. (2019). Why do Latino survey respondents acquiesce? Respondent and interviewer characteristics as determinants of cultural patterns of acquiescence among Latino survey respondents. *Cross-Cultural Research*, 53(1), 87-115.
- Fowler, F. J., & Mangione, T. W. (1989). *Standardized survey interviewing: Minimizing interviewer-related error* (Vol. 18). Thousand Oaks, CA: SAGE Publications, Inc.
- Lee, S., Alvarado-Leiton, F., Yu, W., Davis, R.E., and Johnson, T. Developing a Short Screener for Acquiescent Respondents. *Research in Social and Administrative Pharmacy*. (In press) <https://doi.org/10.1016/j.sapharm.2021.06.012>
- Saris, W. E., Revilla, M., Krosnick, J. A., & Schaeffer, E. M. (2010). Comparing questions with agree/disagree response options to questions with item-specific response options. *Survey Research Methods*, 4(1), 61-79.
- Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, 61(4), 576-602.

Negated and Polar Opposite Items for Balanced Scale  
Construction:  
An Empirical Cross-Cultural Assessment

# Multi-item scale

|                                                             | <b>SD</b>                | <b>D</b>                 | <b>A</b>                 | <b>SA</b>                |
|-------------------------------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <i>Satisfaction with Life</i>                               |                          |                          |                          |                          |
| In most ways my life is close to my ideal                   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| The conditions of my life are excellent                     | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| I am satisfied with my life                                 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| So far, I have gotten the important things I want in life   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| If I could live my life over, I would change almost nothing | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

**SD**= Strongly Disagree, **D**= Disagree, **A**= Agree, **SA**= Strongly Agree

# Unbalanced & Balanced Scales

|                               |                                                             | <b>Unbalanced</b>        |                          |                          |                          | <b>Balanced</b>                                             |
|-------------------------------|-------------------------------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|-------------------------------------------------------------|
|                               |                                                             | <b>SD</b>                | <b>D</b>                 | <b>A</b>                 | <b>SA</b>                |                                                             |
| <i>Satisfaction with Life</i> | In most ways my life is close to my ideal                   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | In most ways my life is not close to my ideal               |
|                               | The conditions of my life are excellent                     | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | The conditions of my life are excellent                     |
|                               | I am satisfied with my life                                 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | I am not satisfied with my life                             |
|                               | So far, I have gotten the important things I want in life   | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | So far, I have gotten the important things I want in life   |
|                               | If I could live my life over, I would change almost nothing | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | If I could live my life over, I would change almost nothing |

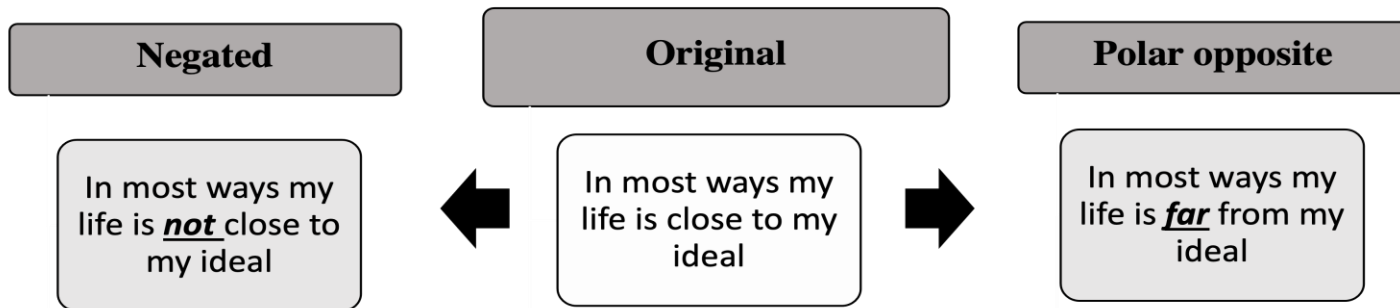
**SD**= Strongly Disagree, **D**= Disagree, **A**= Agree, **SA**= Strongly Agree

# Balanced Scales –a controversial solution

*Why are they controversial?*

- **Reduced reliability** (Schriesheim, Eisenbach & Hill, 1991; Roszkowski & Soven, 2010 )
- **Loss of one-dimensionality** (Benson & Hocevar, 1985; Gnambs & Schroeders, 2017)

# Reversing items



# Justification

There is evidence to support that statement reversal strategies are not equivalent (Schriesheim, Eisenbach & Hill, 1991; Paradis & Willners, 2006).

- Statements with negation particles ("no" or "not") are cognitively more burdensome than other statements (Weems, Onwuegbuzie & Collins, 2006; Sliter & Zickar; 2014; Gnambs & Schroeders, 2017).
- Statements with negation particles ("no" or "not") produce non-ARS-associated illogical answers (Swain, Weathers & Neidrich, 2008; Weijters & Baumgartner, 2012 ).

**Compare two strategies for creating balanced scales in their measurement properties and contrast them with unbalanced wording**



# Hypothesis

- H1: Scales written using polar opposites will yield better measurement properties than scales written using negations.
- H2: Balanced scale wordings (negated or polar opposite) will yield similar CFA and convergent validity measures but lower reliability than unbalanced scales.

# Methodology

Web survey

Questionnaire: Various measures of well-being

*3 samples from online non-probability panels*

Non-Hispanic White respondents in the US interviewed in English (n=1,200)

Hispanic respondents in the US interviewed in English (n=1,200)

Hispanic respondents in Mexico interviewed in Spanish (n=1,200)

# Wording experiment

For four scales respondents were randomly assigned to unbalanced, negated or polar opposite scale wordings.

Satisfaction with Life

Sense of Control

Need for Affect

Social Provisions

Randomization was done independently for each scale.

# Analyses

Reliability –Cronbach's alpha

Factor Structure –One-factor & One-factor + ARS CFA

Convergent validity –Pearson correlations with other constructs

# Participants

|                                                              | <b>Non-Hispanic<br/>White<br/>(n=1,200)</b> | <b>Hispanic<br/>Mex<br/>(n=1,200)</b> | <b>Hispanic US<br/>(n=1,200)</b> |
|--------------------------------------------------------------|---------------------------------------------|---------------------------------------|----------------------------------|
| Age in years M (SD)                                          | 47.8 (18.0)                                 | 42.9 (14.7)                           | 39.7 (15.0)                      |
| Gender (% female)                                            | 49.5                                        | 51.4                                  | 61.7                             |
| Married (%)                                                  | 42.4                                        | 51.1                                  | 52.8                             |
| High School or less (%)                                      | 44.6                                        | 30.1                                  | 46.3                             |
| Use only or mainly Spanish to<br>communicate with family (%) | 1.6                                         | 95.2                                  | 70.0                             |
| Language used to complete the<br>survey                      | English                                     | Spanish                               | English                          |
| ARS- agreement count (SD)                                    | 46.7 (26.4)                                 | 38.6 (21.2)                           | 55.6 (30.9)                      |

# Results-Reliability

*Mean for Cronbach's Alpha across experimental scales*

| <b>Wording</b> | <b>Non-Hispanic<br/>White</b> | <b>Hispanic Mex</b> | <b>Hispanic US</b> |
|----------------|-------------------------------|---------------------|--------------------|
| Unbalanced     | 0.90                          | 0.89                | 0.89               |
| Negated        | 0.73                          | 0.70                | 0.45               |
| Polar Opposite | 0.76                          | 0.72                | 0.46               |

# Results- Factor Structure

*Mean for CFI and RMSEA across experimental scales*

| Model                | Non-Hispanic White |       | Hispanic Mex |       | Hispanic US |       |
|----------------------|--------------------|-------|--------------|-------|-------------|-------|
|                      | CFI                | RMSEA | CFI          | RMSEA | CFI         | RMSEA |
| CFA-- Content only   |                    |       |              |       |             |       |
| Unbalanced           | 0.92               | 0.11  | 0.94         | 0.08  | 0.98        | 0.07  |
| Negated              | 0.72               | 0.19  | 0.67         | 0.18  | 0.61        | 0.24  |
| Polar Opposite       | 0.68               | 0.21  | 0.65         | 0.18  | 0.59        | 0.23  |
| CFA- Content + ARS   |                    |       |              |       |             |       |
| Unbalanced           | NA                 | NA    | NA           | NA    | NA          | NA    |
| Negated + ARS        | 0.94               | 0.07  | 0.94         | 0.05  | 0.99        | 0.04  |
| Polar Opposite + ARS | 0.96               | 0.05  | 0.93         | 0.05  | 0.99        | 0.04  |

# Results-Convergent validity

*Correlations with other constructs: Social Provisions scale*

| Wording        | Non-Hispanic White   |                          | Hispanic Mex         |                          | Hispanic US          |                          |
|----------------|----------------------|--------------------------|----------------------|--------------------------|----------------------|--------------------------|
|                | With Purpose in Life | With Depression Symptoms | With Purpose in Life | With Depression Symptoms | With Purpose in Life | With Depression Symptoms |
| Unbalanced     | 0.46                 | -0.12                    | 0.32                 | -0.03                    | 0.31                 | 0.04                     |
| Negated        | 0.55                 | -0.40*                   | 0.55*                | -0.37*                   | 0.60*                | -0.48*                   |
| Polar Opposite | 0.51                 | -0.48*                   | 0.54*                | -0.39*                   | 0.58*                | -0.42*                   |

\* Significant difference with unbalanced wording ,  $p < 0.05$



# Recap

## *Goal of the study*

Compare two strategies for creating balanced scales in their measurement properties and contrast them with unbalanced wording

# Recap

## *In summary*

*H1: Scales written using polar opposites will yield better measurement properties than scales written using negations.*

The two versions of balanced scales had similar measurement properties

*H2: Balanced scale wordings (negated or polar opposite) will yield similar CFA and convergent validity measures but lower reliability than unbalanced scales.*

- (1) Including an ARS factor into the CFA model improved model fit over unbalanced scales
- (2) Convergent validity was better for balanced scales than unbalanced scales
- (3) Cronbach's alpha was higher for unbalanced scales

# Recap

## *Implications*

How balanced scales are created appears to have little effect on measurement properties of balanced scales.

# References

- Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143–156
- Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement*, 22(3), 231-240.
- Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6(4), 475–494.
- Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, 57, 119–130.
- Gnambs, T., & Schroeders, U. (2017). Cognitive Abilities Explain Wording Effects in the Rosenberg Self-Esteem Scale. *Assessment*, 27(2), 404-418.
- Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological assessment*, 25(4), 1137.
- Roberts, C. (2016). Response styles in surveys: understanding their causes and mitigating their impact on data quality. *The SAGE Handbook of Survey Methodology*.
- Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, 35(1), 113-130.
- Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, 51(1), 67-78.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2), 195-217.

# Response Style and Measurement of Satisfaction with Life

**Sunghee Lee, Wenshan Yu<sup>1</sup>, Jenni Kim<sup>1</sup>,  
Fernanda Alvarado Leiton<sup>1</sup>, Rachel Davis<sup>2</sup>**

<sup>1</sup> University of Michigan, <sup>2</sup> University of South Carolina

Data collection for this study was supported by the Michigan Institute for Clinical & Health Research [grant #: U052427; PI: Lee], the National Institutes of Health/National Cancer Institute [grant number: 5-R01-CA-172283-05; PI: Davis], and the National Science Foundation [grant number SES-1461470, PI: Lee].

# Satisfaction with Life (SWL) Scale – 1

Diener et al. (1985)

**Q2** Please say how much you agree or disagree with the following statements.  
(Mark (X) one box for each line.)

|                                                               | Strongly<br>disagree     | Some<br>what<br>disagree | Slightly<br>disagree     | Neither<br>agree<br>nor<br>disagree | Slightly<br>agree        | Some<br>what<br>agree    | Strongly<br>agree        |
|---------------------------------------------------------------|--------------------------|--------------------------|--------------------------|-------------------------------------|--------------------------|--------------------------|--------------------------|
| In most ways my life is close to ideal.                       | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| The conditions of my life are excellent.                      | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| I am satisfied with my life.                                  | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| So far, I have gotten the important things I want in life.    | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| If I could live my life again, I would change almost nothing. | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
|                                                               | <b>1</b>                 | <b>2</b>                 | <b>3</b>                 | <b>4</b>                            | <b>5</b>                 | <b>6</b>                 | <b>7</b>                 |

- **Scoring norms**

- Sum (responses)
    - >30: Extremely satisfied;
    - >25: Satisfied;
    - >20: Slightly satisfied; ...
    - ≤9: Ext. dissatisfied (see Pavot & Diener, 1993)
  - Mean score

# Satisfaction with Life (SWL) Scale – 2

- Issues
  - All items are written in positive direction
  - Subject to response styles
  - Difficult to ascertain the true satisfaction apart from the effect of response styles (e.g., high SWL score)
- Used widely for cross-cultural, cross-national comparisons
  - Response styles vary by comparison groups

# This Study

- Randomized experiment
  - Directionality of SWL scale items
  - 1 or 2 items re-written in the dissatisfaction direction (Unbalanced vs. Balanced)
  - Randomly split sample
  - Implemented in four different surveys targeting various racial, ethnic and linguistic groups in the US



# Experiment in This Study

- Unbalanced (Control)
  1. In most ways, my life is close to my ideal.
  2. The conditions of my life are excellent.
  3. I am satisfied with my life.
  4. So far, I have gotten the important things I want in life.
  5. If I could live my life over, I would change almost nothing.
- Balanced\* (Experiment)
  1. In most ways, my life is far from my ideal.
  2. The conditions of my life are mediocre.
  3. I am satisfied with my life.
  4. So far, I have gotten the important things I want in life.
  5. If I could live my life over, I would change a lot of things.

\* 2-item balanced in Surveys 1 and 2 (Scale-level randomization);  
1- and 2-item balanced in Survey 3 and 4 (Item-level randomization);  
Balanced in all surveys; Surveys 1-3; Survey 4

# Surveys in This Study

|                               | Survey 1                   | Survey 2                       | Survey 3                                     | Survey 4                                   |
|-------------------------------|----------------------------|--------------------------------|----------------------------------------------|--------------------------------------------|
| <b>Mode</b>                   | Web                        | Telephone                      | Web                                          | Web                                        |
| <b>Target group</b>           | Hispanics                  | Hispanics with<br>ARS tendency | Hispanics, NH<br>Whites, NH Blacks           | Foreign-born<br>Koreans                    |
| <b>Sampling</b>               | River sampling             | List-based quota<br>sampling   | Quota sampling                               | Respondent driven<br>sampling              |
| <b>Sample size</b>            | 1,167                      | 885                            | 2,511                                        | 635                                        |
| <b>SWL Balance</b>            | 0 item: 550<br>1 item: 617 | 0 item: 450<br>1 item: 435     | 0 item: 646<br>1 item: 1,178<br>2 items: 687 | 0 item: 159<br>1 item: 317<br>2 items: 159 |
| <b>Interview<br/>language</b> | Spanish (100%)             | Spanish (93%)<br>English       | Spanish (29%;<br>65%)<br>English             | Korean (67%)<br>English                    |
| <b>Avg. age</b>               | 41 years old               | 64 years old                   | 47 years old                                 | 37 years old                               |
| <b>% Male</b>                 | 45%                        | 20%                            | 46%                                          | 40%                                        |
| <b>% ≤HS edu</b>              | 26%                        | 75%                            | 46%                                          | 13%                                        |

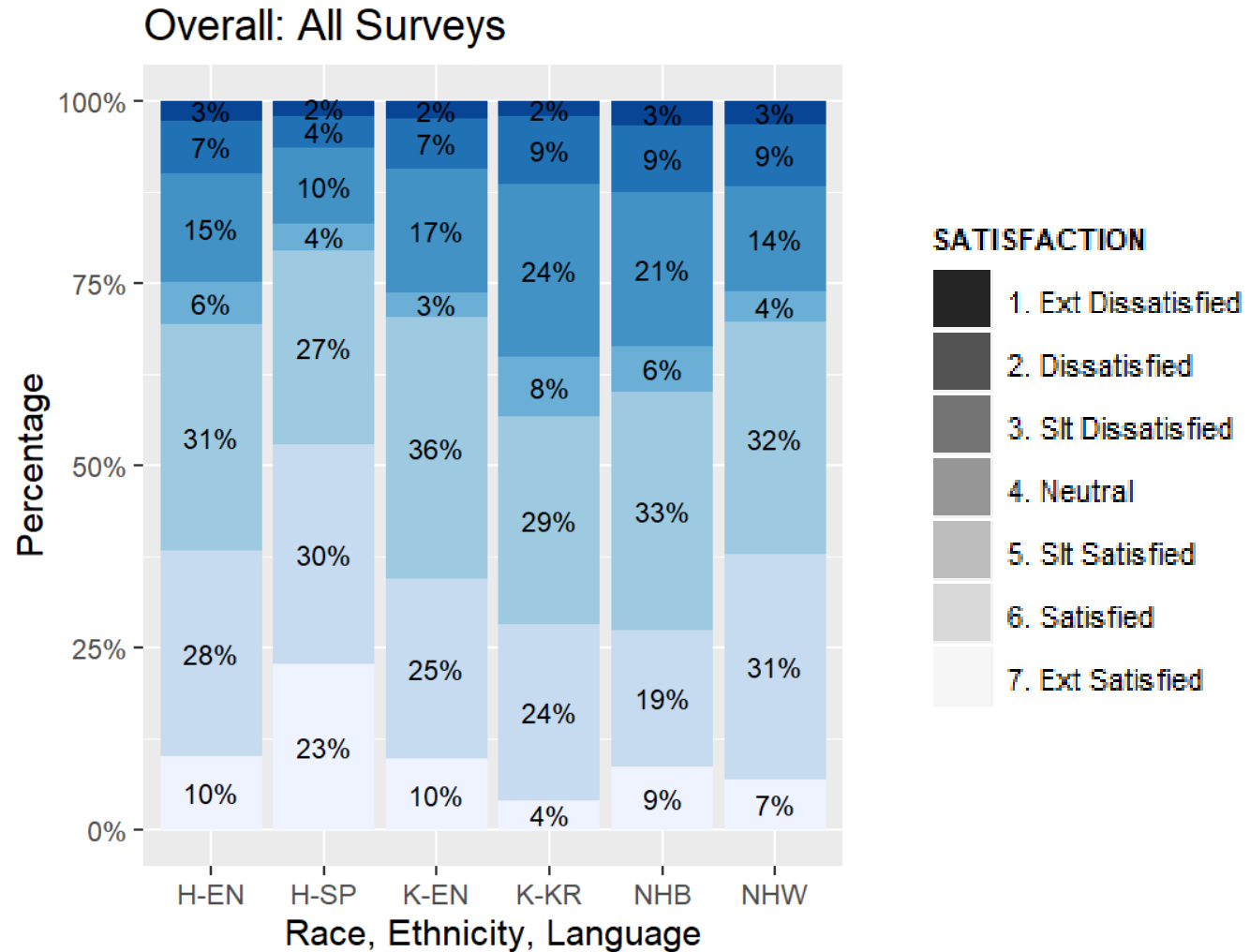
# Analysis

- Responses chosen on SWL items
  - Acquiescent, disacquiescent, extreme and middle responses
- SWL summary scores
  - 1. Extremely satisfied (sum>30) to 7. Extremely dissatisfied (sum≤9)
  - % Extremely satisfied
  - Average score ~ Education + Income
- SWL scale reliability through Cronbach's  $\alpha$
- By race, ethnicity & language groups, by experiment condition, and by survey
  - Groups: H-EN; H-SP; NHW; NHB; K-EN; K-KR

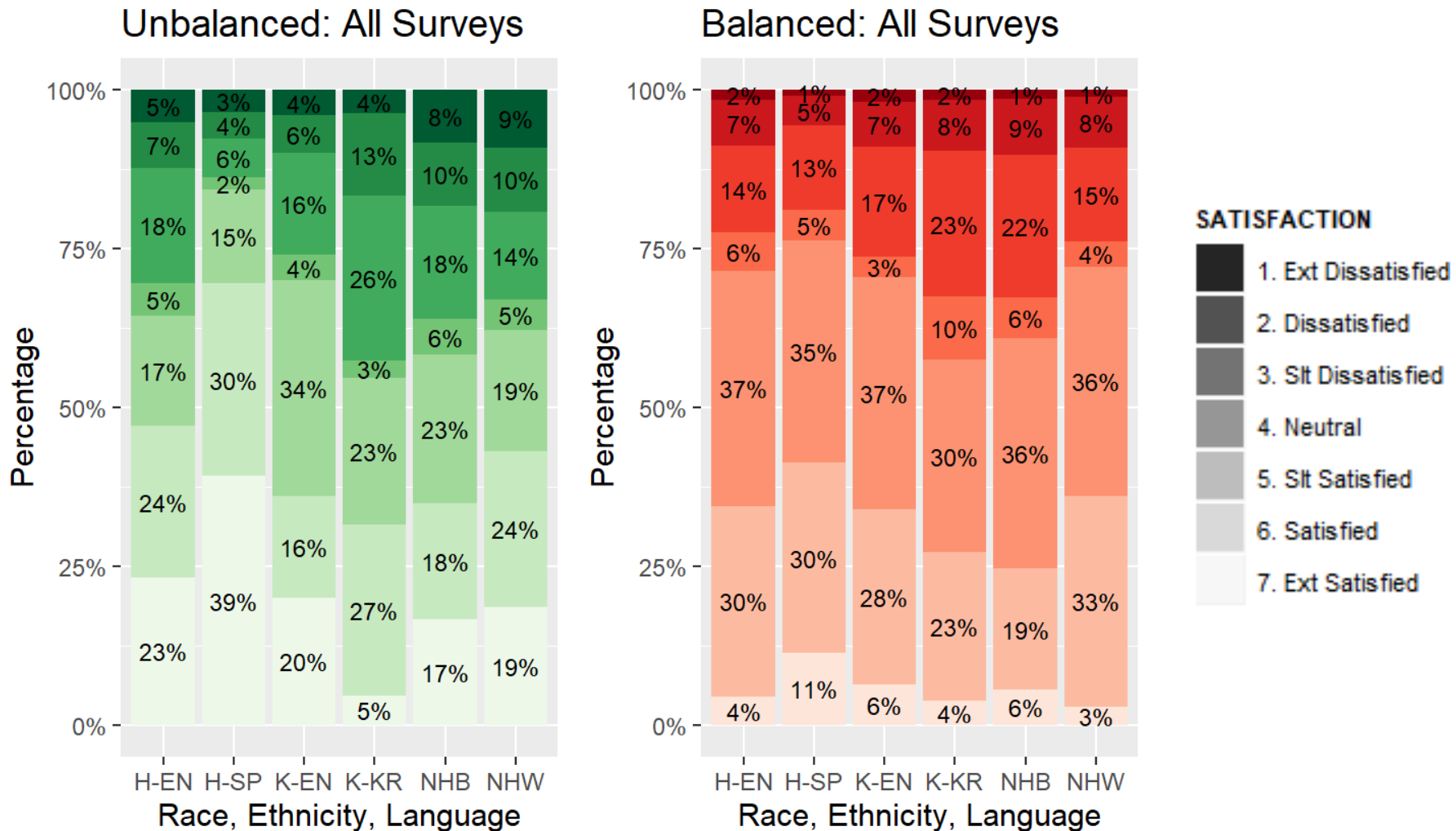
# Responses Chosen on SWL Items

|                 |             | <i>n</i> | Strongly Agree & Agree | Strongly Disagree & Disagree | Strongly (Agree & Disagree) | Neither |
|-----------------|-------------|----------|------------------------|------------------------------|-----------------------------|---------|
| <b>Survey 1</b> | <b>H-SP</b> | 1,237    | 50%                    | 18%                          | 25%                         | 6%      |
| <b>Survey 2</b> | <b>H-EN</b> | 61       | 54%                    | 13%                          | 52%                         | 9%      |
|                 | <b>H-SP</b> | 847      | 71%                    | 11%                          | 67%                         | 4%      |
| <b>Survey 3</b> | <b>H-EN</b> | 393      | 38%                    | 20%                          | 29%                         | 14%     |
|                 | <b>H-SP</b> | 720      | 49%                    | 19%                          | 33%                         | 12%     |
|                 | <b>NHB</b>  | 660      | 29%                    | 26%                          | 28%                         | 14%     |
|                 | <b>NHW</b>  | 738      | 42%                    | 21%                          | 30%                         | 12%     |
| <b>Survey 4</b> | <b>K-EN</b> | 209      | 44%                    | 12%                          | 20%                         | 11%     |
|                 | <b>K-KR</b> | 428      | 31%                    | 14%                          | 10%                         | 23%     |

# Life Satisfaction Distribution

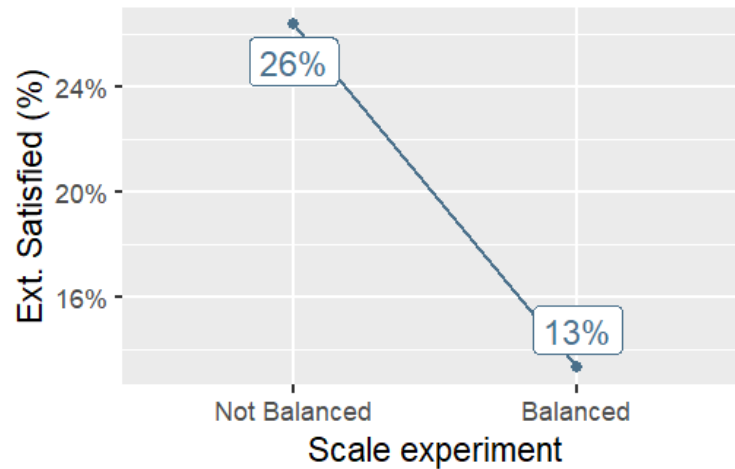


# Satisfaction Distribution by Experiment

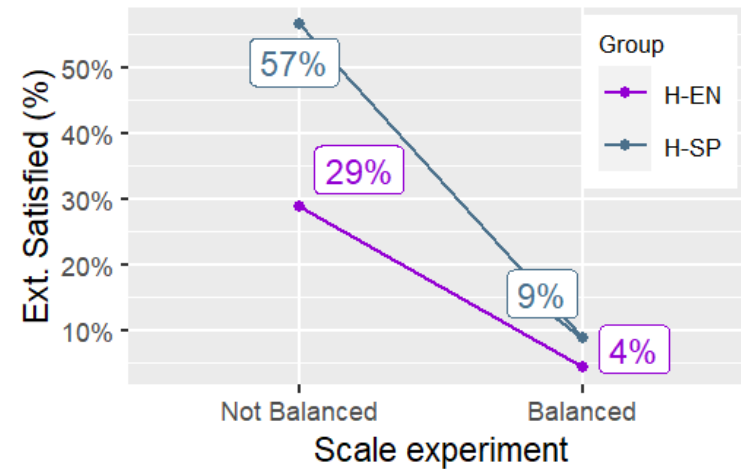


# Extremely Satisfied (%)

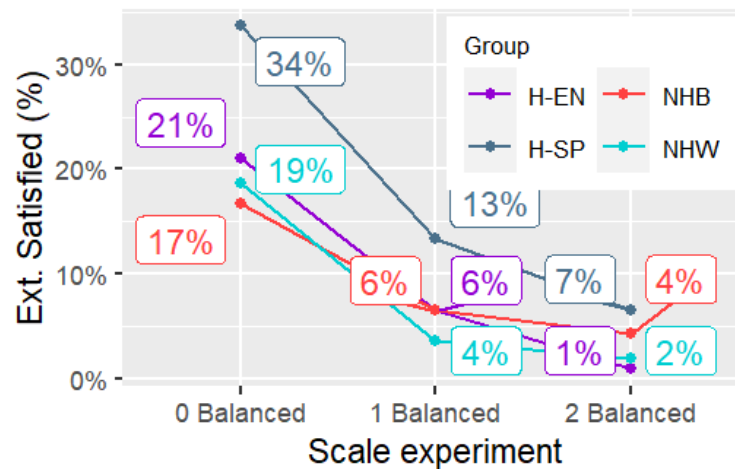
Survey 1: H-SP Only



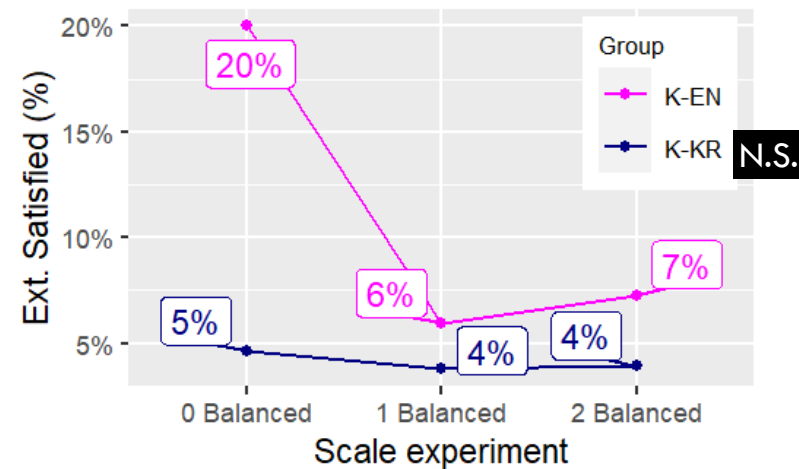
Survey 2



Survey 3



Survey 4



# SWL Score $\sim$ Education + Income\*

|                      | Survey 1 |      | Survey 2<br>(H-SP Only) |      | Survey 4<br>(K-KR Only) |      |      |
|----------------------|----------|------|-------------------------|------|-------------------------|------|------|
| Balanced?            | No       | Yes  | No                      | Yes  | 0                       | 1    | 2    |
| <b>Education</b>     |          |      |                         |      |                         |      |      |
| $\geq$ vs. $<$ HS    |          |      | -0.35                   | 0.21 |                         |      |      |
| $>$ vs. $\leq$ HS    | 0.14     | 0.44 |                         |      |                         |      |      |
| $\geq$ vs. $<$ Col   |          |      |                         |      | 0.24                    | 0.46 | 0.44 |
| <b>Income</b>        |          |      |                         |      |                         |      |      |
| $>$ vs. $\leq$ \$20K |          |      | 0.10                    | 0.02 |                         |      |      |
| $\geq$ vs. $<$ \$50K | 0.30     | 0.34 |                         |      | 0.20                    | 0.36 | 0.36 |

\* Controlled for age and sex.

**Bold**  $p < 0.1$ ; **Bold**  $p < 0.05$



# Cronbach's $\alpha$

|                 |             | Unbalanced |          | Balanced |          |
|-----------------|-------------|------------|----------|----------|----------|
|                 |             | <i>n</i>   | $\alpha$ | <i>n</i> | $\alpha$ |
| <b>Survey 1</b> | <b>H-SP</b> | 624        | 0.85     | 613      | 0.64     |
| <b>Survey 2</b> | <b>H-EN</b> | 38         | 0.80     | 23       | 0.50     |
|                 | <b>H-SP</b> | 412        | 0.83     | 409      | 0.52     |
| <b>Survey 3</b> | <b>H-EN</b> | 100        | 0.90     | 293      | 0.50     |
|                 | <b>H-SP</b> | 178        | 0.90     | 542      | 0.57     |
|                 | <b>NHB</b>  | 180        | 0.88     | 480      | 0.48     |
|                 | <b>NHW</b>  | 188        | 0.93     | 550      | 0.42     |
| <b>Survey 4</b> | <b>K-EN</b> | 50         | 0.89     | 157      | 0.77     |
|                 | <b>K-KR</b> | 108        | 0.90     | 316      | 0.82     |

# Implications (thus far)

- Balanced SWL scale results in
  - Significant changes in response distributions
  - Lower satisfaction scores (% extremely satisfied)
    - Across all surveys
    - Across all racial/ethnic/linguistic groups but K-KR
  - Changes ranking across groups
  - Logical relationship with its correlates (even with K-KR)
  - Lower reliability

# More to come

- Response style vs. Straight-lining
- SWL scale measurement models
  - Unbalanced vs. balanced scale (e.g., CFA)
  - Correcting for response styles (e.g., CFA with response style factor)
- Group comparisons of various SWL scores with and without response style correction
- Expanded concurrent validity
  - Relationship between various SWL scores with validation measures

# References

- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, 49, 71-75.
- Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological Assessment*, 5(2), 164-172.  
<https://doi.org/10.1037/1040-3590.5.2.164>

Thank You!