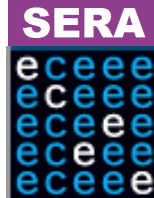# LIKERT SCALES ARE TOO SIMPLISTIC:

Better and More Useful Alternatives in Energy Efficiency

ECEEE 2021 Virtual Summer Study

Dana D'Souza & Lisa Skumatz
Skumatz Economic Research Associates (SERA)
skumatz@serainc.com

PRESENTATION OVERVIEW

SERA – WHO WE ARE

- ◉ Economics Research / Consulting Firm
- ◉ Program Evaluation
- ◉ Been with SERA over 12 years
- ◉ Extensive work in energy and evaluation surveys and validation

OUTLINE

- ◉ Background / Basics of Likert scale
- ◉ Review of Alternatives
- ◉ Applications to energy efficiency evaluation
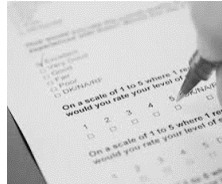- ◉ Summary

Background / Basics

1

LIKERT SCALES AND SURVEY DATA COLLECTION
Common uses: Degrees of….

Importance

Frequency

Agreement

Communiqué PR

Likelihood

Source: Freepik

- Usually 5 point, 7-point, 9-point scales, Neutral center point, unipolar, or bipolar
- Scaled responses of 1, 2, …
- Used in energy surveys for process evaluation, NTG, other
- Likert notes justifiable analysis is only higher / lower, <u>not</u> proportional
- Likert's simplicity seems to encourage misuse / short cut analyses that are not justified

Analysis Errors

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	2	3	4	5
Intensity not linear	Are not equal intervals	Provide ranking order	Not 2x negative as 2	Extreme category avoidance

Looking for More Robust Options

2

What About VAS?

◉ Visual Analog Scale

◉ Non-linear / non-interval

◉ Anchors at the two extremes

◉ Continuous intervals between 1-100

◉ Less end aversion bias than Likert

◉ Potentially improvement and useful, although "numeric" can still be difficult for some respondents

SEARCHING FOR BETTER VALU APPROACHES FOR NEB/NEI

NEB/NEI
- Non Energy Benefits or Impacts
- Measuring and monetizing effects beyond energy savings.

WTP/WTA
- Willingness to pay or willingness to accept.
- Usually first suggestion
- Participant surveys
- Confusion and need for clarifications
- Specific dollar values difficult to provide

LIKERT / VAS
- Considered these approaches.
- Needed more robust method
- Still had analysis issues
- More than directionality

LABELED SCALING
- Labeled Magnitude / Hedonic / Affected / OPUS scaling approaches*
- Relative valuations that can be used as ratios / relatives
- Are not directly numeric
- More easily answered for surveys
- Direct, applied multipliers (& can check within your samples)

*Differences between LS options discussed in paper

Key Points for LS
(labeled scaling)

1) These academic value ratios are well-estimated with confidence intervals and other statistics, and are estimated and tested in multiple studies.

2) SERA's research on NEIs has found that these academic values are extremely consistent with the within-sample LS multipliers.

3) The values are NOT linear; instead, the distances vary *meaningfully* with different labeled modifiers.

4) The work for NEIs requires a few more adaptations, but this is not the focus of this talk*

*See numerous Skumatz / SERA publications including ECEEE 2019, 2021, and prior

MULTIPLIER VALUES

	Labeled Hedonic Scale (LHS)	Labeled Affective Magnitude (LAM)	Oral Pleas- antness and Unpleasantness Scale (OPUS)		Generalized Labeled Magnitude Scaling (g-LMS)
(Like) Greatest Imaginable	**100.00**	**100.00**	**100.00**	**Strongest Imaginable(Like) =>**	**100.00**
Extremely				Very Strong=>	
Very much				Strong=>	
Moderately				Moderate=>	
Slightly				Weak=>	
Neutral	**0.00**	**0.00**	**0.00**	**Neutral=>**	**0.00**
(Dislike) Slightly				Weak=>	
Moderately				Moderate=>	
Very much				Strong=>	
Extremely				Very Strong=>	
(Dislike) Greatest Imaginable	**-100.00**	**-100.00**	**-100.00**	**Strongest Imaginable (dislike)=>**	**-100.00**
Language / Use =>	Like/Dislike	Like/Dislike	Pleasant/Un.		Strength

Source: Skumatz calculations (2021) from multiple sources. Cite if used.
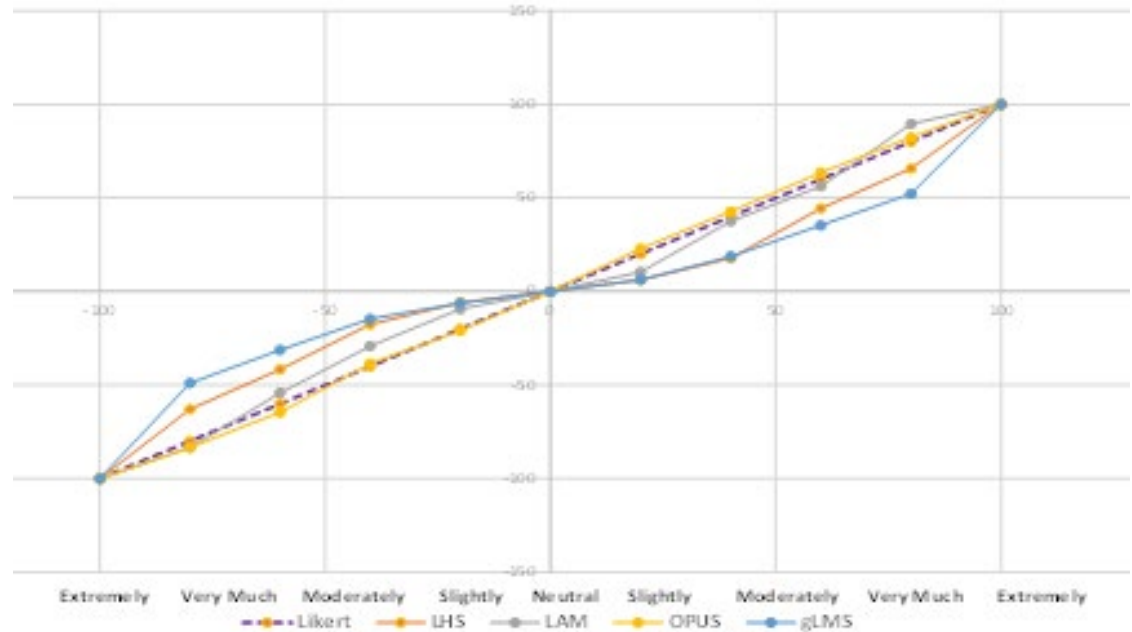
MULTIPLIER VALUES

	Labeled Hedonic Scale (LHS)	Labeled Affective Magnitude (LAM)	Oral Pleas- antness and Unpleasantness Scale (OPUS)		Generalized Labeled Magnitude Scaling (g-LMS)
(Like) Greatest Imaginable				Strongest Imaginable(Like) =>	
Extremely				Very Strong=>	
Very much				Strong=>	
Moderately	*17.82*	*37.50*	*42.71*	*Moderate=>*	*18.75*
Slightly				Weak=>	
Neutral				Neutral=>	
(Dislike) Slightly				Weak=>	
Moderately	*-17.59*	*-29.17*	*-38.54*	*Moderate=>*	*-14.58*
Very much				Strong=>	
Extremely				Very Strong=>	
(Dislike) Greatest Imaginable				Strongest Imaginable (dislike)=>	
Language / Use =>	Like/Dislike	Like/Dislike	Pleasant/Un.		Strength

Source: Skumatz calculations (2021) from multiple sources. Cite if used.
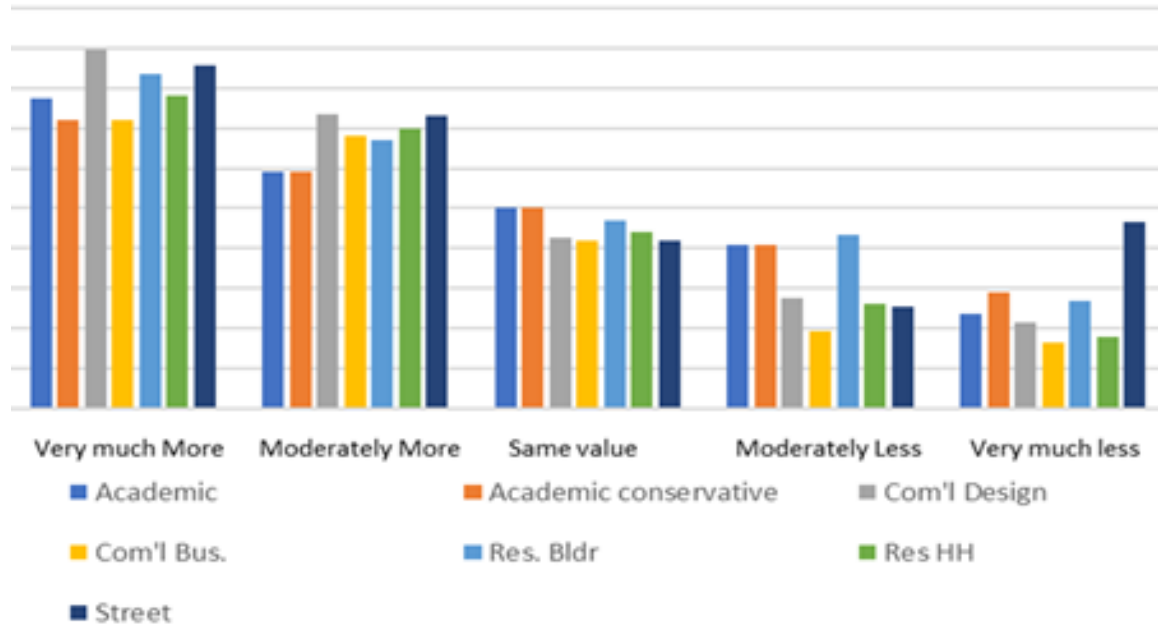
SEMANTIC PHRASES AND SCALE VALUES

FOR MULTIPLE POSITIVE AND NEGATIVE DESCRIPTORS

Non-linear

Comparison of Academic LS multiplier Values with Sample Values for a SERA NEBEl study for Multiple Categories of Interviewees

Fairly close; some Deviations especially for small sample Populations: (street & com'l design).

Very close for other studies with larger samples.

Applications to Energy Efficiency Evaluation 3

LABELED SCALING KEY EM&V APPLICATIONS

- ◉ Process Evalutions

- ◉ Net to Gross / Free-Ridership / Spillover

- ◉ Quantifying Better/ Worse Program Options

- ◉ Barrier Analysis

EXAMPLE 1: BETTER LIKERT SCALES FOR PROCESS EVALUATION SATISFACTION+ QUESTIONS

- Can get more robust / defensible analytical results using same number of response options (5 points each here)

- Allows better representation of differences between categories (difference between 2&3 <u>not</u> same as difference between 1&2), but uses the <u>words</u>

- Multipliers can be scaled to 1, 100, etc; ratios <u>are</u> meaningful & important here

- Useful for satisfaction, likelihood, awareness, many other questions

A. 5-Point Likert Label for "Satisfied with Program Application Process"	B. Common Likert % values	C. Suggested Labeled Scaling Labels for "Satisfied with Program Application Process" - 5 point*	D. g-LMS values	E. LHS values	F. LAM values	G. OPUS values
Very Dissatisfied - 1	0%	Extremely dissatisfied	-0.52	-0.66	-0.90	-0.82
2	25%	Moderately dissatisfied	-0.15	-0.19	-0.29	-0.39
3	50%	Neither satisfied nor dissatisfied	0.00	0.00	0.00	0.00
4	75%	Moderately satisfied	0.19	0.18	0.10	0.23
Very Satisfied – 5	100%	Extremely satisfied	0.52	0.66	0.90	0.82
Table Note: (*) Note that Labeled options support 5, 7, 9-point scales, with special wording.						

EXAMPLE 2: IMPROVING CALCULATIONS THAT CMISUSETLY LIKERT

- Review your existing (survey-based) calculations
- Correct items that implicitly assume "same differences" between survey responses
- Example: MA NTG (multi-step with "corroborating info")
- Steps with potential mis-application of Likert. Corrected figures in paper.
- ➤ *Fixing several steps led to 5.5% difference in NTG estimate**

Example: NTG in one state-calculation steps

Free ridership steps (uses "equal" Likert?)
- Prior intentions – No
- Same Timing of purchase? – Yes
- Same efficiency level?- Yes
- Quantity same? – Yes
- Influence of program – No (10 points)

Spillover
- Screening, ID measures & efficiency– No
- Influence level of program – Yes
- Actions in absence of program partic – Yes
- Consistency check - No

*Thanks to NMR for working with us on this application's recalculation

EXAMPLE RANKING PERFORMANCE
MULTIPLE OPTIONS / POLICIES BETWEEN OPTIONS WORSE

- Traditional scoring of policy program options usually uses H/M/L or Worse, Better / Much better or similar.
  - Usually assume equal differences between values – <u>but</u> move from L to M <u>isn't</u> same as M to H (says LS).

- Revise to Labeled Scaling – provides associated <u>Ratios</u> that incorporate the scoring nuances between these differences.

- Can lead to different conclusions, especially if some options are at extremes in some categories

City Selecting Strategy for GHG Reduction - Energy vs. Recycling-Related								
1 is fastest / best / cheapest	A. Ratio $/MTCE	B. Relative Jobs/$ spent	C. Relative Speed to implement	D. Difficulty for City to Implement (Likert, 1=easy; 5=difficult)	E. Difficulty for City to Implement (LS words)	F. Labeled Scaling score/10	G. Weighted Final Score - Columns A, B, C, D - Using Likert	H. Weighted Final Score - Columns A, B, C, F - Labeled Scale
Commercial Lighting Program	1	1	3	5	Extremely difficult	9	3.4	5.4
LI Weatherization	3	2	3	5	Extremely difficult	9	3.9	5.9
Wind	7	3	10	5	Extremely difficult	9	6.2	8.2
PV	17	11	10	5	Extremely difficult	9	9.0	11.0
Pay As you Throw Trash Rates	0.5	5	0.5	1	Slightly difficult	1	1.2	1.2
Curbside Recycling	0.7	2	1	1	Moderately difficult	3.8	1.0	2.4
Curbside Yard Waste	0.7	0.5	1	1	Moderately difficult	3.8	0.9	2.3
Commercial Recycling	2	1	3	3	Very difficult	5.6	2.6	3.9
Commercial Food Scraps	9	0.5	3	3	Very difficult	5.6	4.0	5.3
Criteria Weights	**0.2**	**0.1**	**0.2**	**0.5**		**0.5**		
(Based on study from 2010; some values no longer accurate)								
Rationale for Extremely difficult for City to implement energy programs - assumes city does not have municipal energy utilty and can only influence through PUC								

18

- Process evaluations commonly use LIKERT scales to score program barriers (1-5, 1-7, etc.) and compute averages using the 1-5 as if it is linear.

- Weaknesses of this approach:
  - Doesn't well-estimate extremes
  - Doesn't provide information on size / meaning of changes (3.2-3.0 means what?)
  - Doesn't provide information on what it would take to mend the issue

- Suggestion 1: Using simple Labeled Scaling would address the first point, and provide meaningful ratios and defensible quantitative scores and comparisons.

- Suggestion 2: Paper suggests expanding on Labeled Scaling by using NEB / NEI techniques (LS-based).
  - Goes beyond defensible scoring to providing DOLLAR valuations – nuanced differences
  - Provides the dollar amount needed to remedy the barrier – guiding program investments / incentives.
  - MUCH more "what next" than change in average LIKERT.

- See paper for detailed steps.

Conclusions and Reccommendations

4

SUMMARY / CONCLUSIONS

◉ Likert is simple, but often misused analytically

◉ Even the misuse is biasing – the underlying relationships are NOT inherently linear (2-1 is not the same as 5-4)

◉ Other simple options are much more analytically robust- recommend <u>labeled scaling</u>

◉ Examples of 4 energy efficiency evaluation applications discussed
  - ◉ Process evaluation surveys (improvements in ➔ importance, aware, agree, likelihood, etc.)
  - ◉ Fixing / improving calculations based off (Likert) survey responses (e.g. a NTG calculation)
  - ◉ Rankings / scoring programs or policies using better / worse, or H/M/L
  - ◉ Barriers analysis approaches that allow comparisons *AND* allow "what next"

◉ Advice – Stop using Likert – EASY and more robust options *(that can be defensibly calculated / compared / ranked)* are readily available for nearly all cases.

THANK YOU!

Contact Us:

Lisa Skumatz 360.261.3069

skumatz@serainc.com

Dana D'Souza

dsouza@serainc.com

www.serainc.com