

# Rating Scales & Ranks

Miela Efraim, Braeden Wong, Grace Ye, Michael Yung



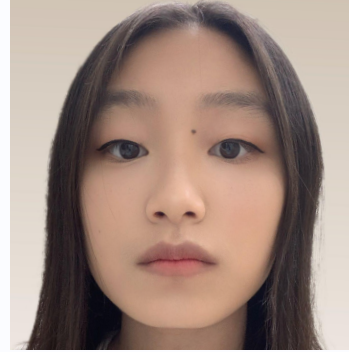
# ... Introductions



**Miela Efraim**  
(she/her)  
MS HFE '24



**Braeden Wong**  
(he/him)  
BS HFE '24



**Grace Ye**  
(she/her)  
BS CS + EC '24



**Michael Yung**  
(he/him)  
BS HFE '24





# Overview of Topic

**Rating Scales:** ask participants to evaluate a product/service by rating their opinion on a predefined scale

- Participants assign a value (sometimes numeric) to the rated object
- Value represents a measure of a specified rated attribute

**Ranking Scales:** ask participants to order items based on specific criteria

- Participants are presented with a list of items
- Asked to rank the items from most to least important
- Process is repeated until all items are compared then ranked



# Qualtrics

- 1st question → Rating scale
- 2nd question → Ranking scale





# Rating Scales

## Uses

- Gather customer feedback
- Measure satisfaction levels
- Identify areas for improvement
- Best used when options are similar, researcher wants to determine degree of difference between them





# Rating Scales

## Advantages

- Easy to use: can be quickly completed
- More reliable data: provides better measure of respondent's attitude/behavior because they can answer more specifically
- Easy to administer and analyze: rating scales produce quantitative data

## Disadvantages

- Limited insights: do not allow respondents to explain their answers in detail
- Limited differentiation: not enough differentiation between answer choices





# Ranking Scales

## Uses

- Identify customer preferences
- Prioritize product features
- Understand importance of different factors

Mobile Phone Features	Rank
Screen Size	
Camera	
Processor	
Battery Life	
Memory	





# Ranking Scales

## Advantages

- Items being compared are NOT similar, and researcher wants to know the most preferred item
- Nuanced insights: discover order of preferences and priorities
- Better differentiation: provide greater distinction between answers

## Disadvantages

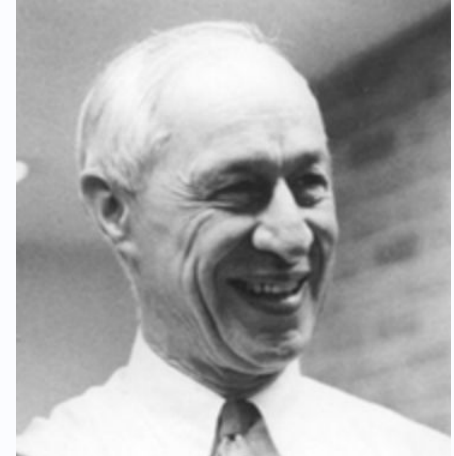
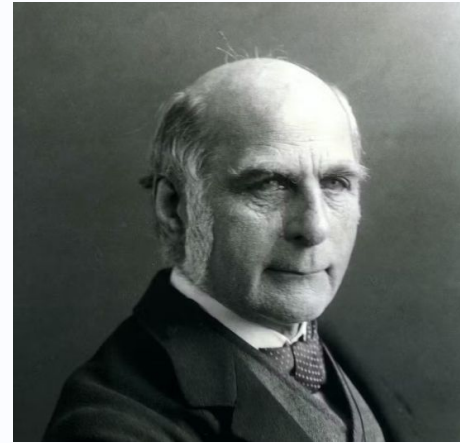
- Hard to identify reasoning: scale may be too small to capture nuance between ratings
- Inaccurate rankings: possibility of respondent's getting lazy





# Background of Topic

- Francis Galton (1879) credited with developing the first rating scale methodologies
  - Five-point scale to describe mental representations of objects (very faint, faint, fair, good, or vivid)
- Rensis Likert contributed to the methodology and application of rating scales in attitude measurement
  - Laid foundation for the widespread use of numeric-label scoring in rating scales, often referred to as "Likert-type" items



# Types of Rating Scales

Likert

Visual  
Analog Scale

Numeric

Semantic  
Differential

Binary



# Binary Scales & Numeric Scales

What does it measure?

- Used to measure **binary outcomes** ("yes/no", "true/false")
- Allows for clear, unambiguous answers
- Allow for rapid data-collection and minimal cognitive effort

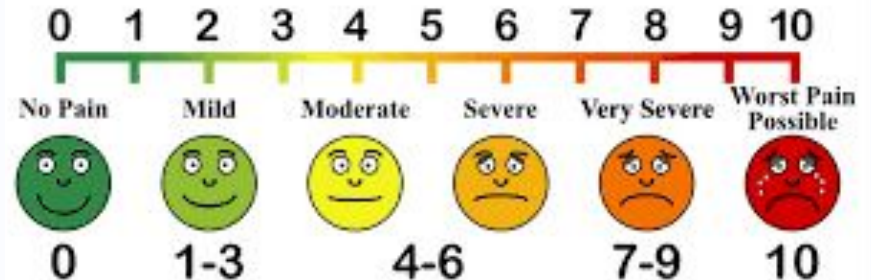
Is this employee ready for promotion?

A Yes

B No

What does it measure?

- Measures preferences, feelings, and perceptions on a numeric scale
- **Assigns a numerical value to quantify responses**
- Most useful when assigning values to subjective parameters





# Likert Scale

What does it measure?

- Measures **level of agreement** to a question or statement
- Follows a 5 or 7 point scale with responses ranging from strongly agree to strongly disagree
- Useful when measuring **intensity** of agreement on a scale

How to interpret the data:

- Measure the **most frequent** responses to understand user sentiment (MODE)
- Best visualized through pie charts or bar charts

## Likert (SUS)

Q1. I think I would like to use this system frequently.

- ☐ Strongly disagree   ☐ Disagree   ☐ Neither agree nor disagree   ☐ Agree   ☐ Strongly agree

Q2. I found the system unnecessarily complex.

- ☐ Strongly disagree   ☐ Disagree   ☐ Neither agree nor disagree   ☐ Agree   ☐ Strongly agree

Q3. I thought the system was easy to use.

- ☐ Strongly disagree   ☐ Disagree   ☐ Neither agree nor disagree   ☐ Agree   ☐ Strongly agree

NNGROUP.COM **NN**/g

## Measure agreement





# Semantic Differential Scale

What does it measure?

- Measures **attitude** between bipolar adjectives
- Generally contains 7 points, but can vary
- Points on the scale are unlabeled which results in a more subjective rating response
- No need to agree with anything but rather choose an option between contrasting words

## Semantic Differential (SEQ)

Overall, this task was:

Very difficult



Very easy

NNGROUP.COM NN/g

Measure **attitude** between  
bipolar adjectives





# Visual Analog Scale (VAS)

What does it measure?

- Measures **intensity** of an attribute on a continuous scale
- Allows for a more detailed analysis
- Subsequently varied interpretation due to the lack of fixed categories
- Most effectively and commonly used when measuring pain or other subjective criteria
- Scoring is determined by measuring the distance of the line between both bipolar values (100 mm)



How hungry are you right now?  
Drag the slider to a point on the scale.

Not at all Extremely

Continue »

The image shows a digital representation of a Visual Analog Scale (VAS). At the top right, there is a green progress bar indicating 40% completion. Below this, the question 'How hungry are you right now?' is displayed, followed by the instruction 'Drag the slider to a point on the scale.' A horizontal slider bar is shown with a green circular handle containing a white right-pointing arrow. The left end of the slider is labeled 'Not at all' and the right end is labeled 'Extremely'. A green button with the text 'Continue »' is located at the bottom right of the slider interface.

Think semantic differential  
but with a slider instead





# System Usability Scale (SUS)

- Industry standard to measure user satisfaction
- Participant feedback is manipulated into a score from 0-100
- Anything below 68 is considered below average and anything above 68 is above average

The System Usability Scale Standard Version		Strongly Disagree					Strongly Agree				
			1	2	3	4	5				
1	I think that I would like to use this system frequently.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
2	I found the system unnecessarily complex.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
3	I thought the system was easy to use.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
4	I think that I would need the support of a technical person to be able to use this system.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
5	I found the various functions in this system were well integrated.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
6	I thought there was too much inconsistency in this system.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
7	I would imagine that most people would learn to use this system very quickly.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
8	I found the system very awkward to use.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
9	I felt very confident using the system.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
10	I needed to learn a lot of things before I could get going with this system.		<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				

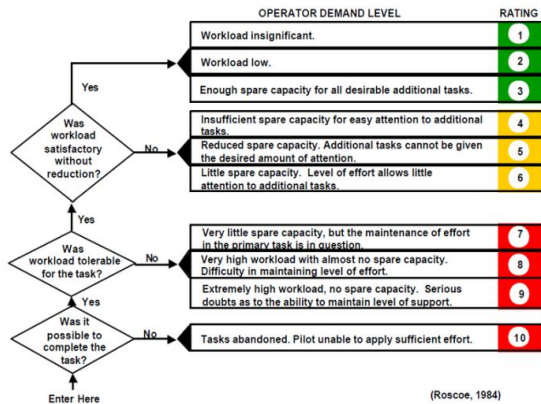


# Bedford Scale + NASA Task Load Index (TLX)

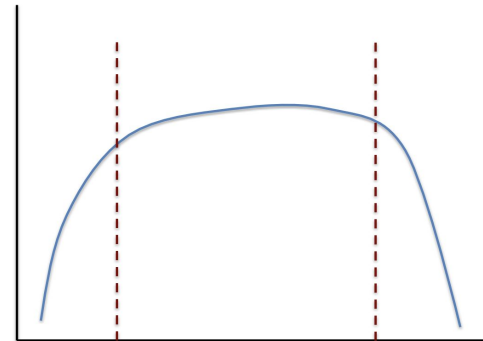
The Bedford Scale and NASA TLX both measure the workload of a task so the designer can evaluate the workload for optimal performance.

## Bedford Scale

- Unidimensional rating scale
- Hierarchical decision tree leading to a 10 point scale



Performance



Workload





# Bedford Scale + NASA Task Load Index (TLX)

The Bedford Scale and NASA TLX both measure the workload of a task so the designer can evaluate the workload for optimal performance.

## TLX

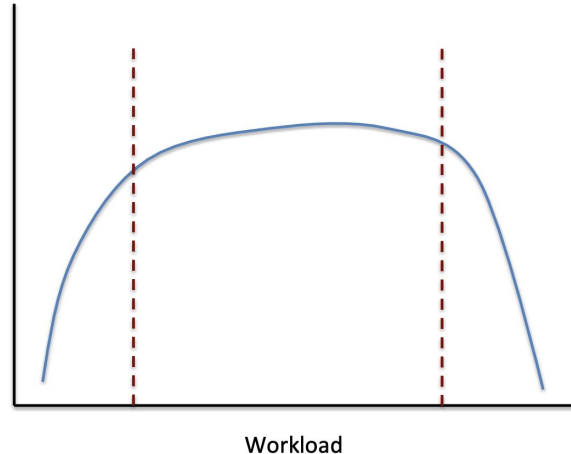
- Multidimensional (6 dimensions) rating scale
- Rating contributions of each dimension to workload to determine intensity of workload

### NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date
<b>Mental Demand</b> How mentally demanding was the task?		
Very Low  Very High		
<b>Physical Demand</b> How physically demanding was the task?		
Very Low  Very High		
<b>Temporal Demand</b> How hurried or rushed was the pace of the task?		
Very Low  Very High		
<b>Performance</b> How successful were you in accomplishing what you were asked to do?		
Perfect  Failure		
<b>Effort</b> How hard did you have to work to accomplish your level of performance?		
Very Low  Very High		
<b>Frustration</b> How insecure, discouraged, irritated, stressed, and annoyed were you?		
Very Low  Very High		

Performance





# Net Promoter Score (NPS)

- Metric for customer loyalty and satisfaction
- How likely are you to recommend our [product] to a friend or colleague? (on a scale from 1-10)
- $NPS = \% \text{ of Promoters} - \% \text{ of Detractors}$
- Common type of **Rating Scale**

**We appreciate your feedback!** X

Thank you for visiting our website. We are always looking for ways to improve your experience. Please take a moment to tell us about your experience.

How likely are you to recommend our website to a friend or colleague?

0 1 2 3 4 5 6 7 8 9 10

What could we do to improve your experience?

[Send Feedback](#)

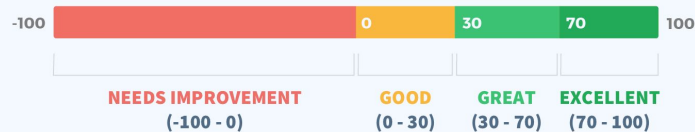
powered by QuestionPro

## Net Promoter Score®

$$= \begin{matrix} \text{😊} \\ \% \\ \text{PROMOTERS} \end{matrix} - \begin{matrix} \text{😞} \\ \% \\ \text{DETRACTORS} \end{matrix}$$



## What is a good NPS score?

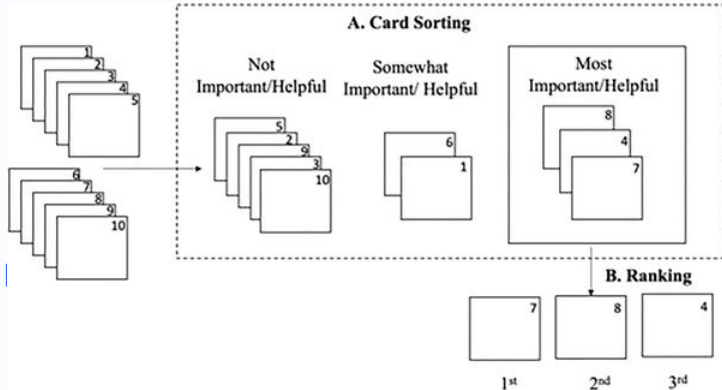




# Ranking Scales

Ranking scales center around the premise of arranging items or options in a specific order based on their perceived value. Examples include:

## Card sorting



## Ordinal Ranking

Mobile Phone Features	Rank
Screen Size	
Camera	
Processor	
Battery Life	
Memory	



# Research Paper: Background

## Overview:

- Rapid growth in the wearable market
- Extensive evaluative testing on the reliability and function of fitness tracking wearables
- Concerns over such products' usability
- Quantitative user reviews of various mainstream fitness trackers
  - Evaluate the perceived usability of various mainstream fitness trackers
  - Receive user feedback on product features

Jia, Y., Wang, W., Wen, D., Liang, L., Gao, L., & Lei, J. (2018). Perceived user preferences and usability evaluation of mainstream wearable devices for health monitoring. *PeerJ (San Francisco, CA)*, 6, e5350–e5350. <https://doi.org/10.7717/peerj.5350>





# Research Paper: Methods

7 fitness trackers (smart wristbands and smart watches) reflective of the Chinese market



Apple Watch



Samsung Gear S



Xiaomi Mi Band



HUAWEI Talkband B2



Fitbit Surge



Jawbone Up3



Misfit Shine





# Research Paper: Methods

388 total participants recruited from researchers' social media (WeChat) and public posters





# ..... Research Paper: Questionnaire

Two-part questionnaire for product feature preference and device usability referencing existing methods for product evaluation for achieve better validity

## **Five-Point Likert Scale**

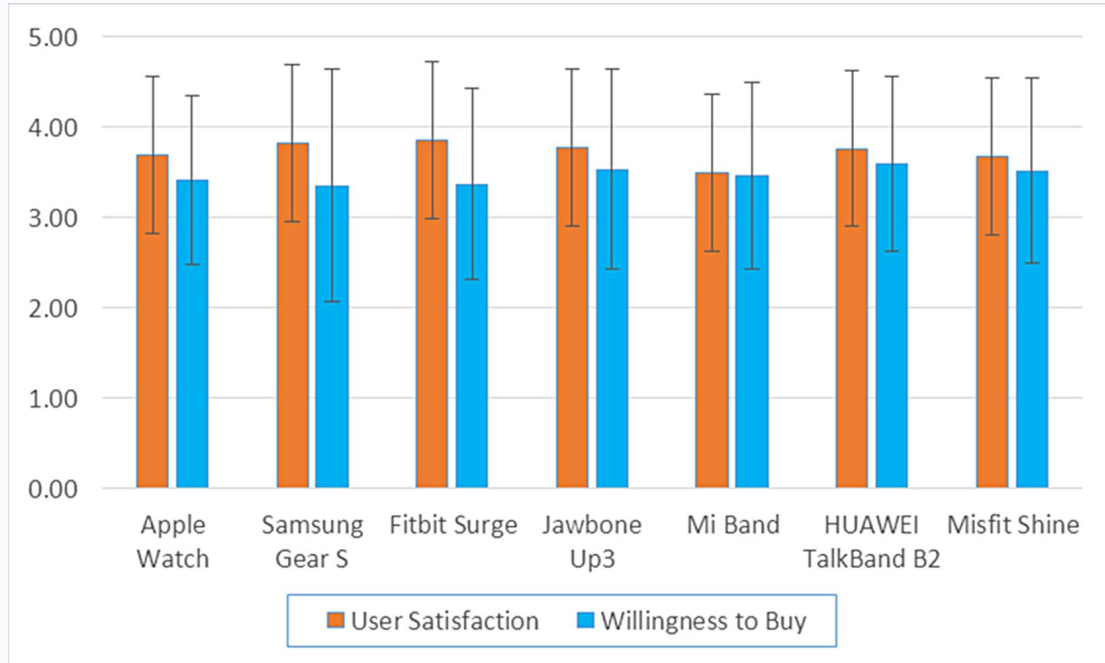
1 strongly dissatisfied, 2 dissatisfied, 3 neutral, 4 satisfied, 5 strongly satisfied

- Product design
- Durability
- Ease of use
- Added features
- And user-rated accuracy





# Research Paper: Results



**Figure 1** The mean satisfaction and willingness to buy scores for each device.





# Research Paper: Results



(A) Apple Watch



(B) Samsung Gear S



(E) Mi Band



(F) Huawei Talkband B2



(C) Fitbit Surge



(D) Jawbone Up3



(G) Misfit Shine

**Figure 2** Radar graphs showing variations of five dimension scores of the 7 Devices.



# Research Paper: Results

	<i>F</i> (6, 382)	<i>p</i> -value
Product design	3.001	0.007
Durability	2.824	0.011
Ease of use	0.431	0.858
Added features	2.076	0.055
Accuracy	1.132	0.343

**Figure 3** Analysis of variance of the scores for each device.

- Significant difference observed in product design and durability among different devices ( $p < 0.05$ )
- No significant differences observed in ease of use, added features, and accuracy among different devices ( $p > 0.05$ )





# .... Research Paper: Conclusion

Subjective positive intent regarding fitness trackers, BUT unsatisfied with their cost effectiveness (less willing to purchase)

Fair ratings of fitness trackers with some significant differences among devices, SO further improvement needed to existing fitness trackers

## Reflections:

- Rating scales can be used to quantitatively differentiate and garner user feedback on products/product iterations for usability and other typically qualitative evaluations
- Should be used throughout the product development life cycle





# ... R Studio Demonstration

- Binary Scale
- 5 Point Scale
- Ranking Scale

