Measurement Scales and Scaling

Dr. Hemal Pandya

Measurement & Scaling

- Assignment of numbers to characteristics of objects, persons, states or events, according to rules
- **Measurement** means assigning numbers or other symbols to characteristics of objects according to certain pre-specified rules.
  - One-to-one correspondence between the numbers and the characteristics being measured.
  - The rules for assigning numbers should be standardized and applied uniformly.
  - Rules must not change over objects or time.

  Scaling: Scaling is an extension of measurement. Scaling involves creating a continuum on which measurements on objects are located.

Importance of measurement

- research conclusions are only as good as the data on which they are based
- observations must be quantifiable in order to subject them to statistical analysis
- the dependent variable(s) must be measured in any quantitative study.
- the more precise, sensitive the method of measurement, the better.

Keys to Measurement

- You do not measure the object, person, state or event, but characteristics of the object
- Numbers are used to represent the observable/unobservable characteristics
- Rules specify how the numbers are to be assigned to the characteristics

Measurement Definitions

- Concepts: Invented name for the property of an object, person, state or event
- Construct: A concept, having been deliberately and consciously invented or adapted for a special scientific purpose
  - Constitutive Def'n
  - Operational Def'n

Constructs and Measurement

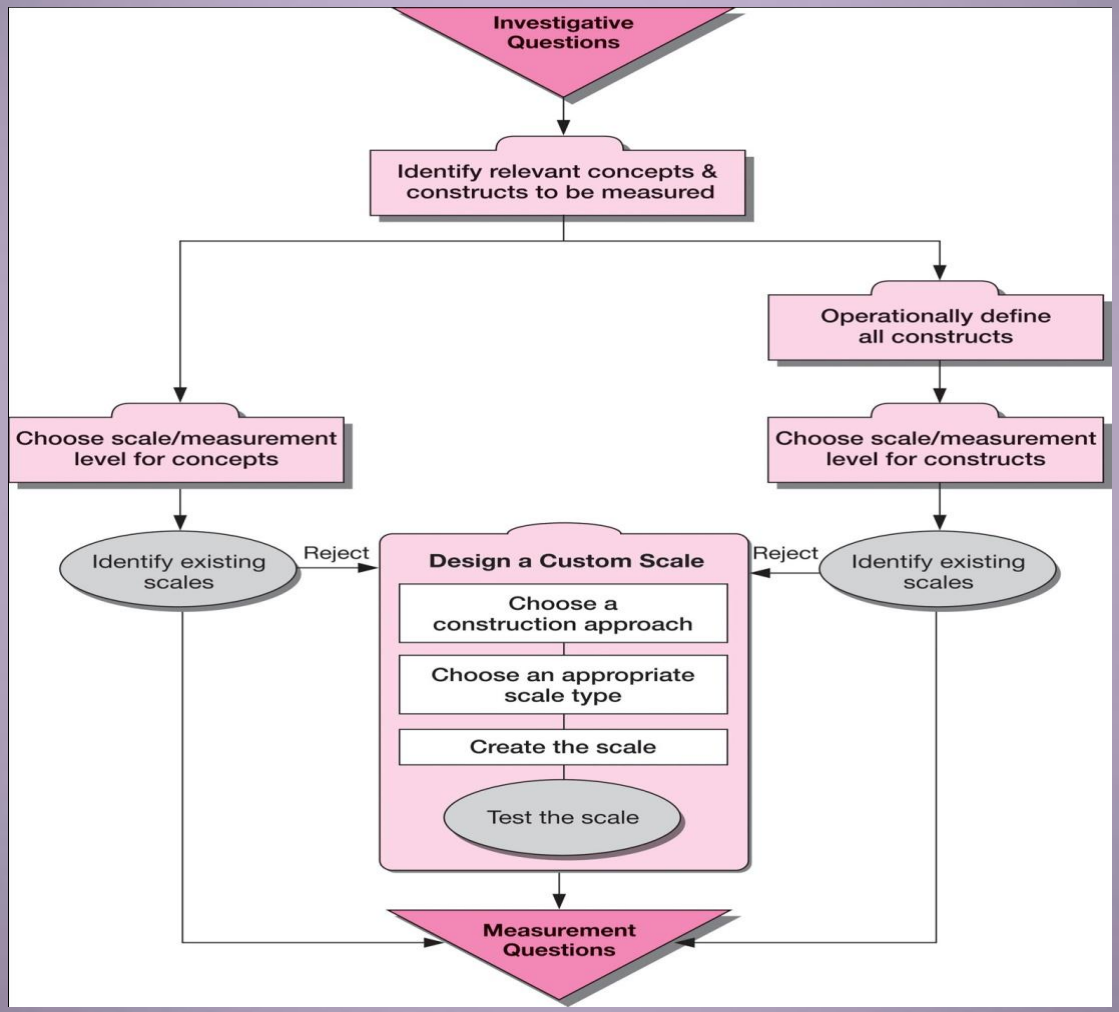
- **Construct Development**
  - Identifying and defining what is to be measured
  - A construct is a hypothetical variable composed of different elements that are thought to be related (e.g., 5 questions tapping brand loyalty)

- **Measurement**
  - Figuring out how to measure what you want to measure
  - A Measure needs to be reliable and valid →

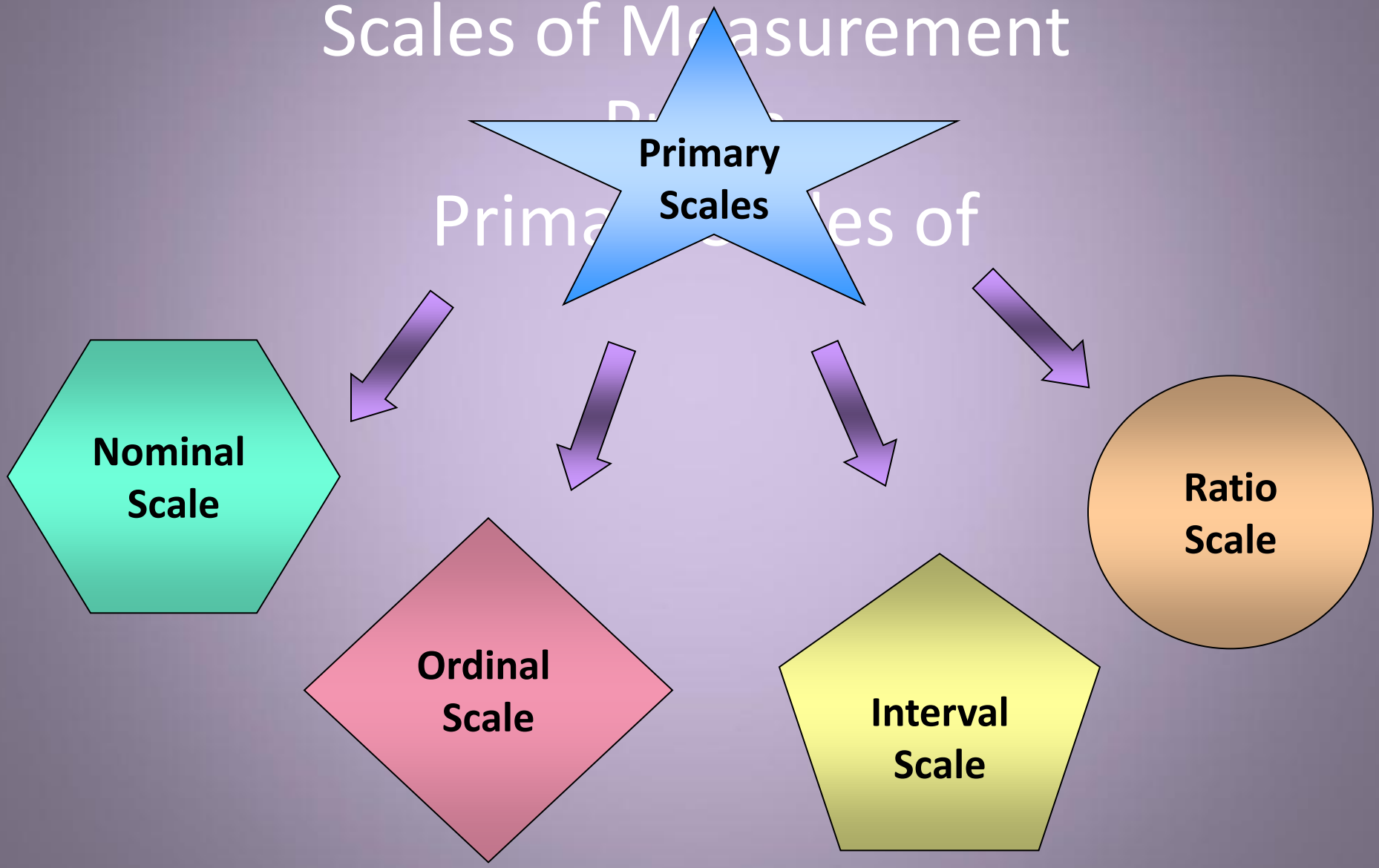
Components of Measurement

- M = C + E
- M = Measurement
- C = Characteristic being measured
- E = Errors

The Scaling Process

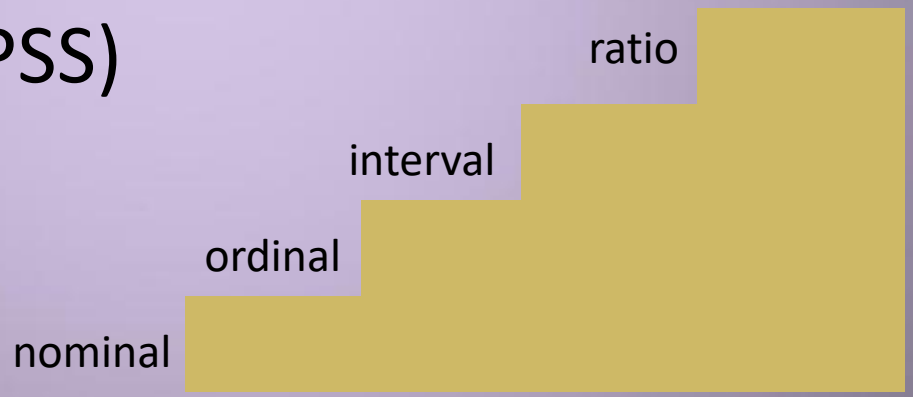


Primary Scales of Measurement



Levels of data

- Nominal
- Ordinal
- Interval (Scale in SPSS)
- Ratio (Scale in SPSS)



Nominal

- The numbers serve only as labels or tags for identifying and classifying objects.
- When used for identification, there is a strict one-to-one correspondence between the numbers and the objects.
- The numbers do not reflect the amount of the characteristic possessed by the objects.
- The only permissible operation on the numbers in a nominal scale is counting.
- Only a limited number of statistics, all of which are based on frequency counts, are permissible, e.g., percentages, and mode.

Nominal – Categorical Scale example

What is your gender?

___ Male ___Female

Ordinal

- A ranking scale in which numbers are assigned to objects to indicate the relative extent to which the objects possess some characteristic.
- Can determine whether an object has more or less of a characteristic than some other object, but not how much more or less.
- Any series of numbers can be assigned that preserves the ordered relationships between the objects.
- In addition to the counting operation allowable for nominal scale data, ordinal scales permit the use of statistics based on centiles, e.g., percentile, quartile, median.

Ordinal Scale Example –

Please rank order the following as to how often you recycle each item where 1=item you most often recycle, 7=item you recycle the least.

___Cardboard ___Glass

___Newspaper ___Plastic

___Other Paper Products ___Aluminum

___Other, please specify_____

Interval

- Numerically equal distances on the scale represent equal values in the characteristic being measured.
- It permits comparison of the differences between objects.
- The location of the zero point is not fixed. Both the zero point and the units of measurement are arbitrary.
- Any positive linear transformation of the form y = a + bx will preserve the properties of the scale.
- It is not meaningful to take ratios of scale values.
- Statistical techniques that may be used include all of those that can be applied to nominal and ordinal data in addition the arithmetic mean, standard deviation, and other statistics commonly used in marketing research.
- **Interval Scales** - it is possible to compare differences in magnitude, but importantly the zero point does *not* have a natural meaning. It captures the properties of nominal and ordinal scales -- used by most psychological tests.

- Designates an equal-interval ordering - The distance between, for example, a 1 and a 2 is the same as the distance between a 4 and a 5
- ***Interval scales:***
  - **Size of difference is known**
  - **Units are of equal size**

- Example - Celsius temperature is an interval variable. It is meaningful to say that 25 degrees Celsius is 3 degrees hotter than 22 degrees Celsius, and that 17 degrees Celsius is the same amount hotter (3 degrees) than 14 degrees Celsius. Notice, however, that 0 degrees Celsius does not have a natural meaning. That is, 0 degrees Celsius does not mean the absence of heat!

Ratio

- Possesses all the properties of the nominal, ordinal, and interval scales.
- It has an absolute zero point.
- It is meaningful to compute ratios of scale values.
- Only proportionate transformations of the form y = bx, where b is a positive constant, are allowed.
- All statistical techniques can be applied to ratio data.
- In addition to all the properties of nominal, ordinal, and interval measures, ratio measures have a true zero point
  - Eg. Length of time
  - Eg. Number of times
  - Eg. Number of affiliations
- Can actually state ratio of one to another
  - X has twice as many affiliations as Y
- ***Ratio scales:***
  - **True zero point exists**
  - **Multiplication or division possible**

Illustration of Primary Scales of Measurement

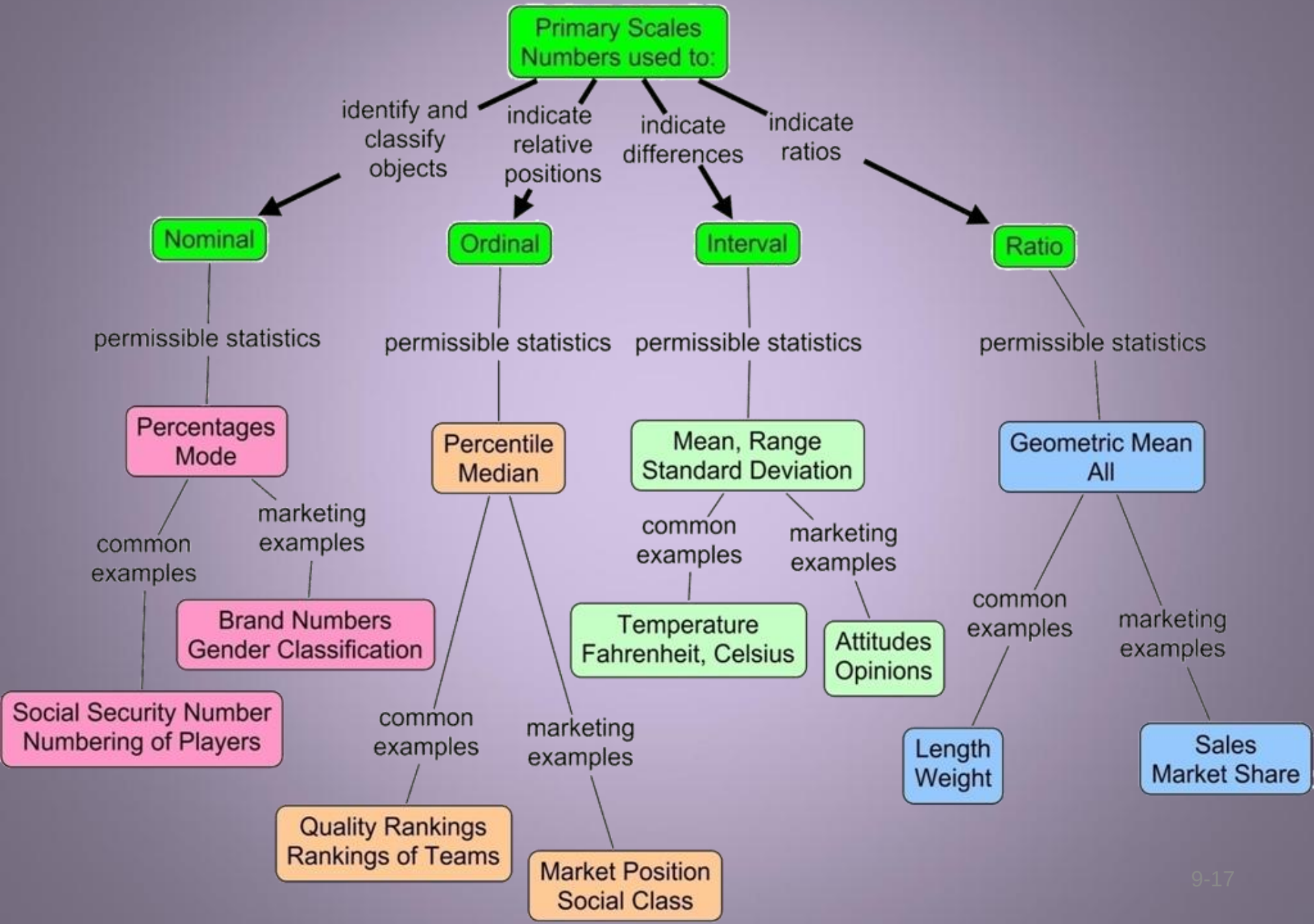
Table 8.2

Nominal Scale	Ordinal Scale	Interval Scale		Ratio Scale
No. Store	Preference Rankings	Preference Ratings		$ spent last 3 months
		1-7	11-17	
1. Lord & Taylor	7	79	5	15	0
2. Macy's	2	25	7	17	200
3. Kmart	8	82	4	14	0
4. Rich's	3	30	6	16	100
5. J.C. Penney	1	10	7	17	250
6. Neiman Marcus	5	53	5	15	35
7. Target	9	95	4	14	0
8. Saks Fifth Avenue	6	61	5	15	100
9. Sears	4	45	6	16	0
10.Wal-Mart	10	115	2	12	10

Types of scales

- Nominal scales--qualitative, not quantitative distinction (no absolute zero, not equal intervals, not magnitude)
- Ordinal scales--ranking individuals (magnitude, but not equal intervals or absolute zero)
- Interval scales--scales that have magnitude and equal intervals but not absolute zero
- Ratio scales--have magnitude, equal intervals, and absolute zero (so can compute ratios)

Figure 9.10 A Concept Map for Primary Scales



Primary scales of measurement

Scale	Basic characteristics	Common examples	Marketing examples	Permissible statistics
Nominal	Numbers identify and classify objects.	Identification Card, numbering of football players.	Brands numbers, store types, sex classification	Percentages, mode
Ordinal	Numbers indicate the relative positions of the objects but not the magnitude of differences between them.	Quality rankings, ranking of teams in a tournament	Preference ranking, market position, social class	Percentile, median
Interval	Differences between objects can be compared; zero point is arbitrary	Temperature (Fahrenheit, Celsius)	Attitudes, opinions, index numbers	Range, mean, standard deviation
Ratio	Zero point is fixed; rations of scale values can be computed	Length, weight	Age, income, costs, sales, market shares.	Geometric mean (All)

Source: Maholtra, N.K. (2004) Marketing Research: An Applied Orientation. Upper-Saddle River, New Jersey, USA: Pearson Education Inc.

Primary Scales of Measurement

Table 8.1

Scale	Basic Characteristics	Common Examples	Marketing Examples	Permissible Statistics	
				Descriptive	Inferential
Nominal	Numbers identify & classify objects	Social Security nos., numbering of football players	Brand nos., store types	Percentages, mode	Chi-square, binomial test
Ordinal	Nos. indicate the relative positions of objects but not the magnitude of differences between them	Quality rankings, rankings of teams in a tournament	Preference rankings, market position, social class	Percentile, median	Rank-order correlation, Friedman ANOVA
Interval	Differences between objects	Temperature (Fahrenheit)	Attitudes, opinions, index	Range, mean, standard	Product-moment
Ratio	Zero point is fixed, ratios of scale values can be compared	Length, weight	Age, sales, income, costs	Geometric mean, harmonic mean	Coefficient of variation

Measurement/Scaling Properties

- **Assignment**
  - You can assign objects to categories
- **Order (Magnitude)**
  - You can order objects in terms of having more or less of some quality
- **Distance (Equal Intervals)**
  - The distance between adjacent points on the scale is identical
- **Origin (Absolute Zero Point)**
  - Zero “means something” (absence of a given quality)

Comparison of Measurement Scales

	Label	Order	Distance	Origin
Nominal scale	Yes	No	No	No
Ordinal scale	Yes	Yes	No	No
Interval scale	Yes	Yes	Yes	No
Ratio scale	Yes	Yes	Yes	Yes

Another way to describe variables

- ***Qualitative variables:*** **have a nominal or ordinal scale of measurement.**
- ***Continuous variables:*** **have an Ordinal, interval, or ratio variables scale of measurement.**
- ***Quantitative variables:*** **have an interval scale of measurement.**
- ***Categorical variables:*** **have a nominal or ordinal scale of measurement.**

What Type of Scale?

- **Number of Sweaters Purchased This Year? _____**
- **What is Your Ethnicity?**
- **To what extent do you agree or disagree that Congress should have approved the $700 bailout? (1 = strongly disagree to 7 = strongly agree)**
- **Please rank the following issues from most to least important (Iraq, Health Care, Economy, Environment)**
- **What is your income? (5-10k; 11-15k; 16-20k; 20-25k; 25-30k)**
- **How many hours have you completed toward your degree?**

  ___under 30 hours ___30-59 hours

  ___60-89 hours ___90 or more hours

Test Your Knowledge:

A professor is interested in the relationship between the number of times students are absent from class and the letter grade that students receive on the final exam. He records the number of absences for each student, as well as the letter grade (A,B,C,D,F) each student earns on the final exam. In this example, what is the measurement scale for number of absences?

a) Nominal b) Ordinal c) Interval d) Ratio

In the previous example, what is the measurement scale of letter grade on the final exam?

a) Nominal b) Ordinal c) Interval d) Ratio

A researcher is interested in studying the effect of room temperature in degrees Fahrenheit on productivity of automobile assembly workers. She controls the temperature of the three manufacturing facilities, such that employees in one facility work in a room temperature of 60 degrees, employees in another facility work in a room temperature of 65 degrees, and the last group works in a room temperature of 70 degrees. The productivity of each group is indicated by the number of automobiles produced each day. In this example, what is the measurement scale of room temperature?

a) Nominal b) Ordinal c) Interval d)Ratio

In the previous example, what is the measurement scale of productivity?

a) Nominal b) Ordinal c) Interval d) Ratio

Select the highest appropriate level of measurement:

Educational Level:

1 = Some High school
2 =High school Diploma
3 = Undergraduate Degree
4 = Masters Degree
5 = Doctorate Degree

a) Nominal b) Ordinal c) Interval d) Ratio

Select the highest appropriate level of measurement:

Number of questions asked during a class lecture

a) Nominal b) Ordinal c) Interval d) Ratio

Select the highest level of measurement:

Categories on a Likert-type scale measuring attitudes:

1 = Strongly Disagree
2 = Disagree
3 = Neutral
4 = Agree
5 = Strongly Agree

a) Nominal b) Ordinal c) Interval d) Ratio

Attitude

An enduring disposition to consistently respond in a given matter

Attitudes as Hypothetical Constructs

- The term hypothetical construct is used to describe a variable that is not directly observable, but is measurable by an indirect means such as verbal expression or overt behavior - attitudes are considered to be such variables.

Three Components of an Attitude

- Affective
- Cognitive
- Behavioral

Three Components of Attitudes

The ABCs of attitudes:

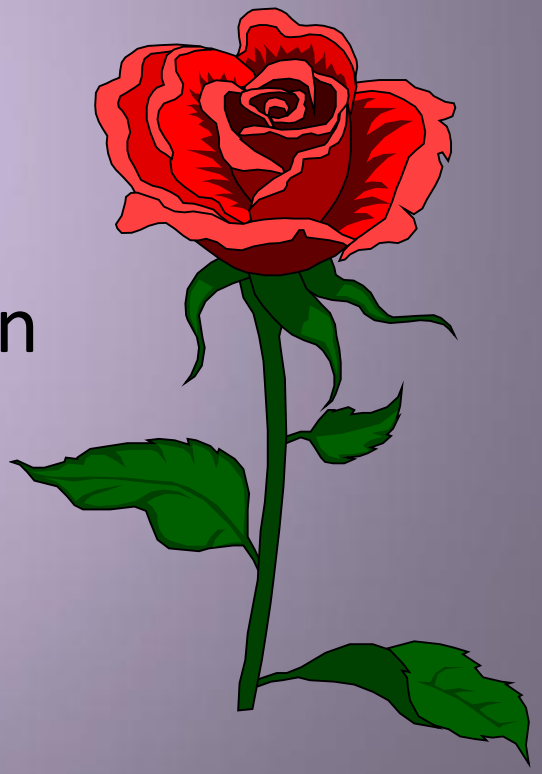
- **The <u>Affective</u> Component (based on feelings or overall evaluation)** Feelings of like or dislike

- **The <u>Behavioral</u> Component (likely action toward object; e.g. from a consumer behavior point of view, the consumer's intention to buy a product)** Intentions to behave

- **The <u>Cognitive</u> Component (based on beliefs; what you think about a marketing stimulus) –** Information possessed

Affective

The feelings or emotions toward an object



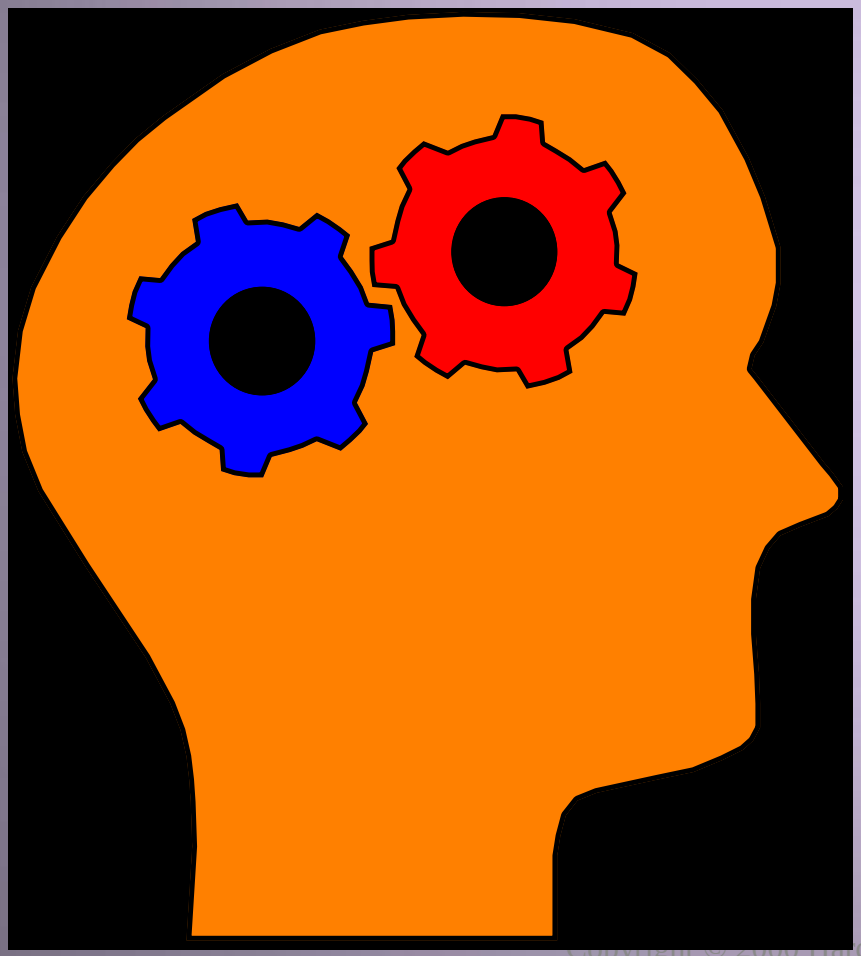
Behavioral

- Predisposition to action
- Intentions
- Behavioral expectations

Cognitive

Knowledge and beliefs

Measuring Attitudes



- Ranking
- Rating
- Sorting
- Choice

Copyright © 2000 Harcourt, Inc. All rights reserved.

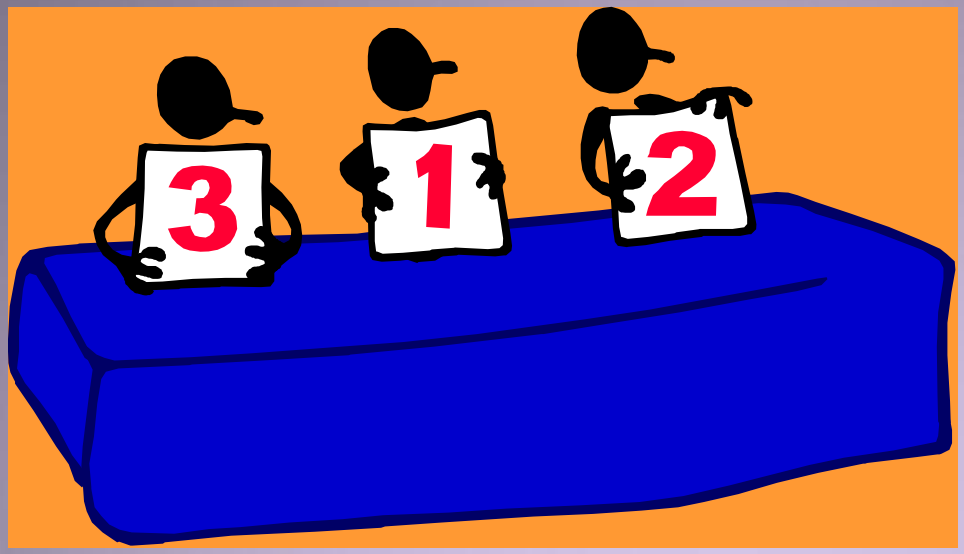
The Attitude Measuring Process

Ranking - Rank order preference

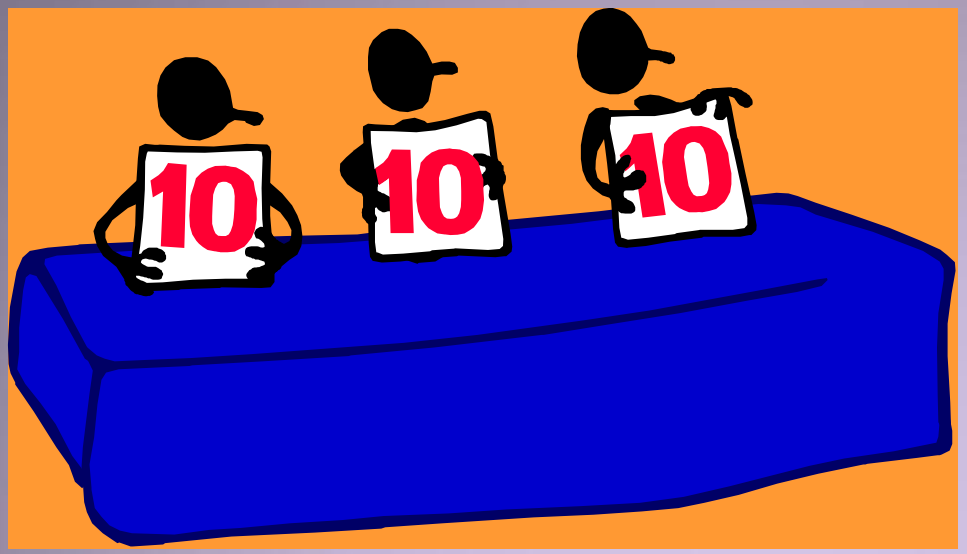
Rating - Estimates magnitude of a characteristic

Sorting - Arrange or classify concepts

Choice - Selection of preferred alternative



- Ranking tasks require that the respondent rank order a small number of objects in overall performance on the basis of some characteristic or stimulus.



- Rating asks the respondent to estimate the magnitude of a characteristic, or quality, that an object possesses. The respondent's position on a scale(s) is where he or she would rate an object.

- Sorting might present the respondent with several concepts typed on cards and require that the respondent arrange the cards into a number of piles or otherwise classify the concepts.

- Choice between two or more alternatives is another type of attitude measurement - it is assumed that the chosen object is preferred over the other.

- Physiological measures of attitudes provide a means of measuring attitudes without verbally questioning the respondent. for example, galvanic skin responses, measure blood pressure etc.

Simple Attitude Scaling

- In its most basic form, attitude scaling requires that an individual agree with a statement or respond to a single question. This type of self-rating scale merely classifies respondents into one of two categories;
- Comparative Scales
- Non-Comparative Scales

Types of Scaling Techniques

- **COMPARATIVE SCALES**
  - **Involve the respondent directly comparing stimulus objects.**
  - **e.g. How does Pepsi compare with Coke on sweetness**

- **NONCOMPARATIVE SCALES**
  - **Respondent scales each stimulus object independently of other objects**
  - **e.g. How would you rate the sweetness of Pepsi on a scale of 1 to 10**

A Classification of Scaling Techniques

Figure 8.2

A Comparison of Scaling Techniques

- **Comparative scales** involve the direct comparison of stimulus objects. Comparative scale data must be interpreted in relative terms and have only ordinal or rank order properties.

- In **non-comparative scales**, each object is scaled independently of the others in the stimulus set. The resulting data are generally assumed to be interval or ratio scaled.

Relative Advantages of Comparative Scales

- Small differences between stimulus objects can be detected.
- Same known reference points for all respondents.
- Easily understood and can be applied.
- Involve fewer theoretical assumptions.
- Tend to reduce halo or carryover effects from one judgment to another.

Relative Disadvantages of Comparative Scales

- Ordinal nature of the data
- Inability to generalize beyond the stimulus objects scaled.

Classification of Scales

Formats of Comparative Scales –

- Paired comparison scales
- Rank order scale
- Constant sum rating scale
- Q-sort technique

- **Non-Comparative Scales –** In the non-comparative scales, the respondents do not make use of any frame of reference before answering the questions.

Comparative Scaling Techniques
Paired Comparison Scaling

- A respondent is presented with two objects and asked to select one according to some criterion.
- The data obtained are ordinal in nature.
- Paired comparison scaling is the most widely used comparative scaling technique.
- With n brands, [n(n - 1) /2] paired comparisons are required
- Under the assumption of transitivity, it is possible to convert paired comparison data to a rank order.

Comparative Scale

- **Comparative scales –** In comparative scales it is assumed that respondents make use of a standard frame of reference before answering the question.

Example:-

- Please rate Domino's in comparison to Pizza Hut on the basis of your satisfaction level on an 11-point scale, based on the following parameters: (1 = Extremely poor, 6 = Average, 11 = Extremely good). Circle your response:

a.	Variety of menu options	1	2	3	4	5	6	7	8	9	10	11
b.	Value for money	1	2	3	4	5	6	7	8	9	10	11
c.	Speed of service (delivery time)	1	2	3	4	5	6	7	8	9	10	11
d.	Promotional offers	1	2	3	4	5	6	7	8	9	10	11
e.	Food quality	1	2	3	4	5	6	7	8	9	10	11
f.	Brand name	1	2	3	4	5	6	7	8	9	10	11
g.	Quality of service	1	2	3	4	5	6	7	8	9	10	11
h.	Convenience in terms of takeaway location	1	2	3	4	5	6	7	8	9	10	11
i.	Friendliness of the salesperson on the phone	1	2	3	4	5	6	7	8	9	10	11
j.	Quality of packaging	1	2	3	4	5	6	7	8	9	10	11
k.	Adaptation of Indian taste	1	2	3	4	5	6	7	8	9	10	11
l.	Side orders/appetizers	1	2	3	4	5	6	7	8	9	10	11

Obtaining Shampoo Preferences Using Paired Comparisons

Figure 8.3

Instructions: We are going to present you with ten pairs of shampoo brands. For each pair, please indicate which one of the two brands of shampoo you would prefer for personal use.

Recording Form:

	Jhirmack	Finesse	Vidal Sassoon	Head & Shoulders	Pert
Jhirmack		0	0	1	0
Finesse	1[a]		0	1	0
Vidal Sassoon	1	1		1	1
Head & Shoulders	0	0	0		0
Pert	1	1	0	1	
Number of Times Preferred[b]	3	2	0	4	1

[a]A 1 in a particular box means that the brand in that column was preferred over the brand in the corresponding row. A 0 means that the row brand was preferred over the column brand. [b]The number of times a brand was preferred is obtained by summing the 1s in each column.

Paired Comparison Items

If we have brands A, B, C and D, we would have respondents compare

- A and B
- A and C
- A and D
- B and C
- B and D
- C and D

–Usually limited to N < 15

COMPARATIVE SCALES

Paired Comparison

Please indicate which of the following airlines you prefer by circling your more preferred airline in each pair:

Air Canada	WestJet
Air Transat	Air Canada
Zip	WestJet
WestJet	Air Transat
Air Canada	Zip
Zip	Air Transat

Comparative Scaling Techniques
Rank Order Scaling

- Respondents are presented with several objects simultaneously and asked to order or rank them according to some criterion.
- It is possible that the respondent may dislike the brand ranked 1 in an absolute sense.
- Furthermore, rank order scaling also results in ordinal data.
- Only (n - 1) scaling decisions need be made in rank order scaling.

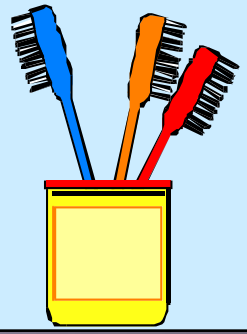
Preference for Toothpaste Brands Using Rank Order Scaling

Figure 8.4

Instructions: Rank the various brands of toothpaste in order of preference. Begin by picking out the one brand that you like most and assign it a number 1. Then find the second most preferred brand and assign it a number 2. Continue this procedure until you have ranked all the brands of toothpaste in order of preference. The least preferred brand should be assigned a rank of 10.

No two brands should receive the same rank number.

The criterion of preference is entirely up to you. There is no right or wrong answer. Just try to be consistent.

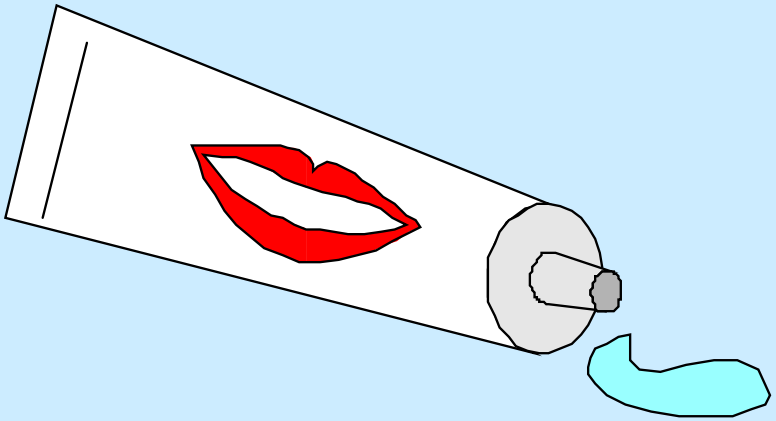


Preference for Toothpaste Brands Using Rank Order Scaling

Figure 8.4 cont.

Form

Brand	Rank Order
1. Crest	_____
2. Colgate	_____
3. Aim	_____
4. Gleem	_____
5. Macleans	_____
6. Ultra Brite	_____
7. Close Up	_____
8. Pepsodent	_____
9. Plus White	_____
10. Stripe	_____



COMPARATIVE SCALES

Rank-Order Scales

Rank the following soft-drinks from 1 (best) to 5 (worst) according to your taste preference:

Coca-Cola _____

7-Up _____

Dr. Pepper _____

Pepsi-Cola _____

Mountain Dew _____

- **Top and bottom rank choices are 'easy'**
- **Middle ranks are usually most 'difficult'**

Comparative Scales

Rank Order Scale

Indicate your preferred type of music with a 1, your second favorite with a 2, and so on for each type of music:

____ **Heavy Metal**

____ **Alternative**

____ **Urban Contemporary**

____ **Classical**

____ **Country**

COMPARATIVE SCALES

Compared to Chevrolet, Ford is:

less innovative			about the same			more innovative
☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7

Comparative Scaling Techniques
Constant Sum Scaling

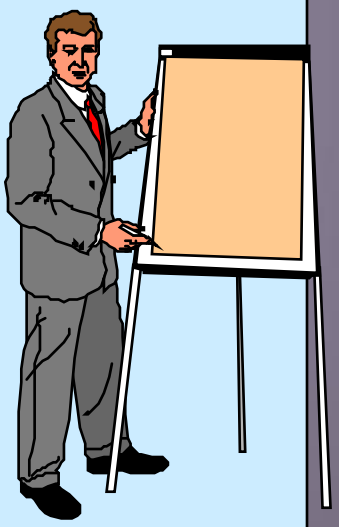
- Respondents allocate a constant sum of units, such as 100 points to attributes of a product to reflect their importance.
- If an attribute is unimportant, the respondent assigns it zero points.
- If an attribute is twice as important as some other attribute, it receives twice as many points.
- The sum of all the points is 100. Hence, the name of the scale.

Importance of Bathing Soap Attributes Using a Constant Sum Scale

Figure 8.5

Instructions

On the next slide, there are eight attributes of bathing soaps. Please allocate 100 points among the attributes so that your allocation reflects the relative importance you attach to each attribute. The more points an attribute receives, the more important the attribute is. If an attribute is not at all important, assign it zero points. If an attribute is twice as important as some other attribute, it should receive twice as many points.



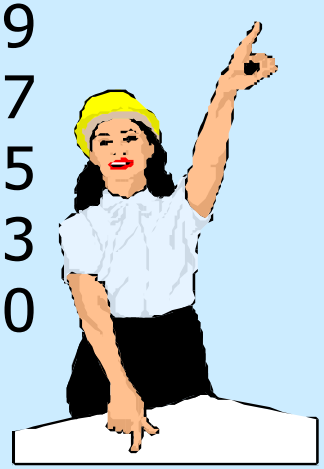
Importance of Bathing Soap Attributes Using a Constant Sum Scale

Figure 8.5 cont.

Form

Average Responses of Three Segments

Attribute	Segment I	Segment II	Segment III
1. Mildness	8	2	4
2. Lather	2	4	17
3. Shrinkage	3	9	7
4. Price	53	17	9
5. Fragrance	9	0	19
6. Packaging	7	5	9
7. Moisturizing	5	3	20
8. Cleaning Power	13	60	15
Sum	100	100	100



COMPARATIVE SCALES

Constant Sum Scales

Allocate a total of 100 points among the following soft-drinks depending on how favorable you feel toward each; the more highly you think of each soft-drink, the more points you should allocate to it. (Please check that the allocated points add to 100.)

Coca-Cola	**_____**	**points**
7-Up	**_____**	**points**
Dr. Pepper	**_____**	**points**
Tab	**_____**	**points**
Pepsi-Cola	**_____**	**points**

100 points

Constant Sum Scale

Please divide 100 points among the following characteristics so the division reflects the relative importance of each characteristic to you in the selection of a bank

Hours of service _____

Friendliness _____

Distance from home _____

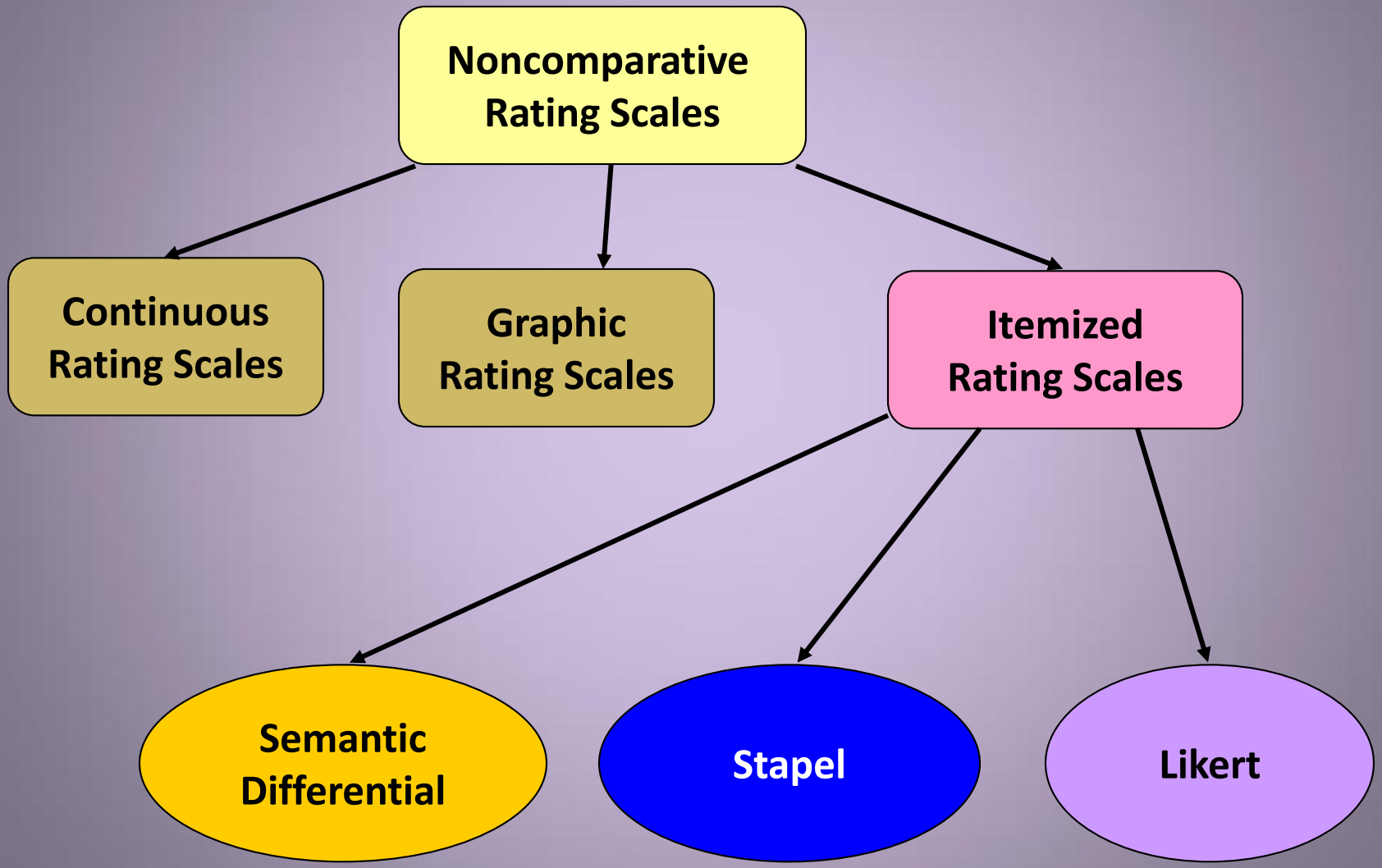
Investment vehicles _____

Parking facilities _____

Q-Sort

- Q -Sort technique uses the rank order procedure in which the objects (statements) are sorted into different piles based on their similarity with respect to certain criterion, by the respondents. The data generated in this way would be ordinal in nature. The distribution of the number of statements (objects) in each pile should be such that the resulting data may follow a normal distribution. The number of piles can be around 10 or more as the large number increases the reliability or precision of the results.

Figure 10.3 A Classification of Noncomparative Rating Scales



Non-comparative Scale : **Continuous Rating Scale**

Respondents rate the objects by placing a mark at the appropriate position on a line that runs from one extreme of the criterion variable to the other. The form of the continuous scale may vary considerably.

How would you rate Sears as a department store?

Version 1

Probably the worst - - - - - - -I - Probably the best

Version 2

Probably the worst - - - - - - -I - Probably the best

0 10 20 30 40 50 60 70 80 90 100

Version 3

Very bad Neither good nor bad Very good

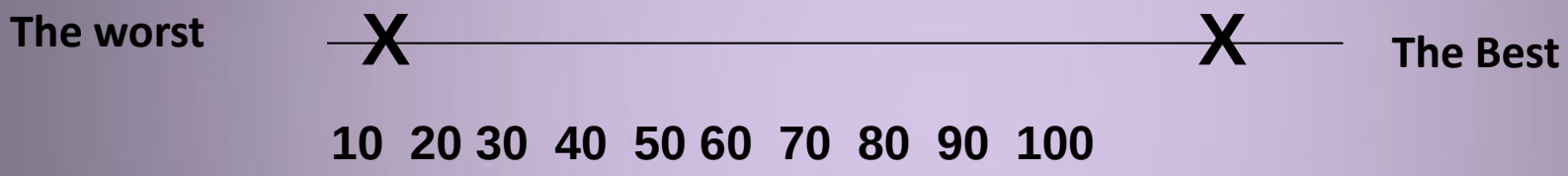
Probably the worst - - - - - - -I - Probably the best

0 10 20 30 40 50 60 70 80 90 100

Non comparative scale

Continuous scale

- How would you rate Marketing Research to other courses this term

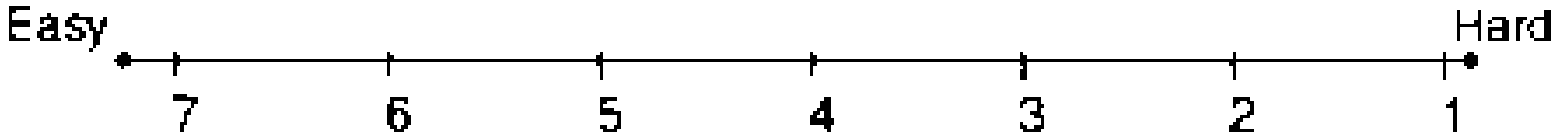


How easy or difficult did you find Acme pasta to cook?

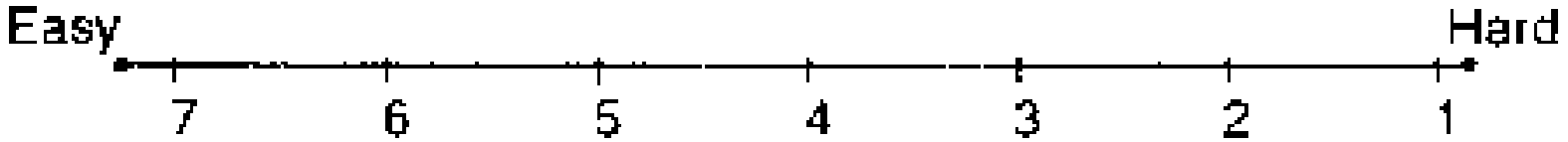
Type A

Type B

Type C



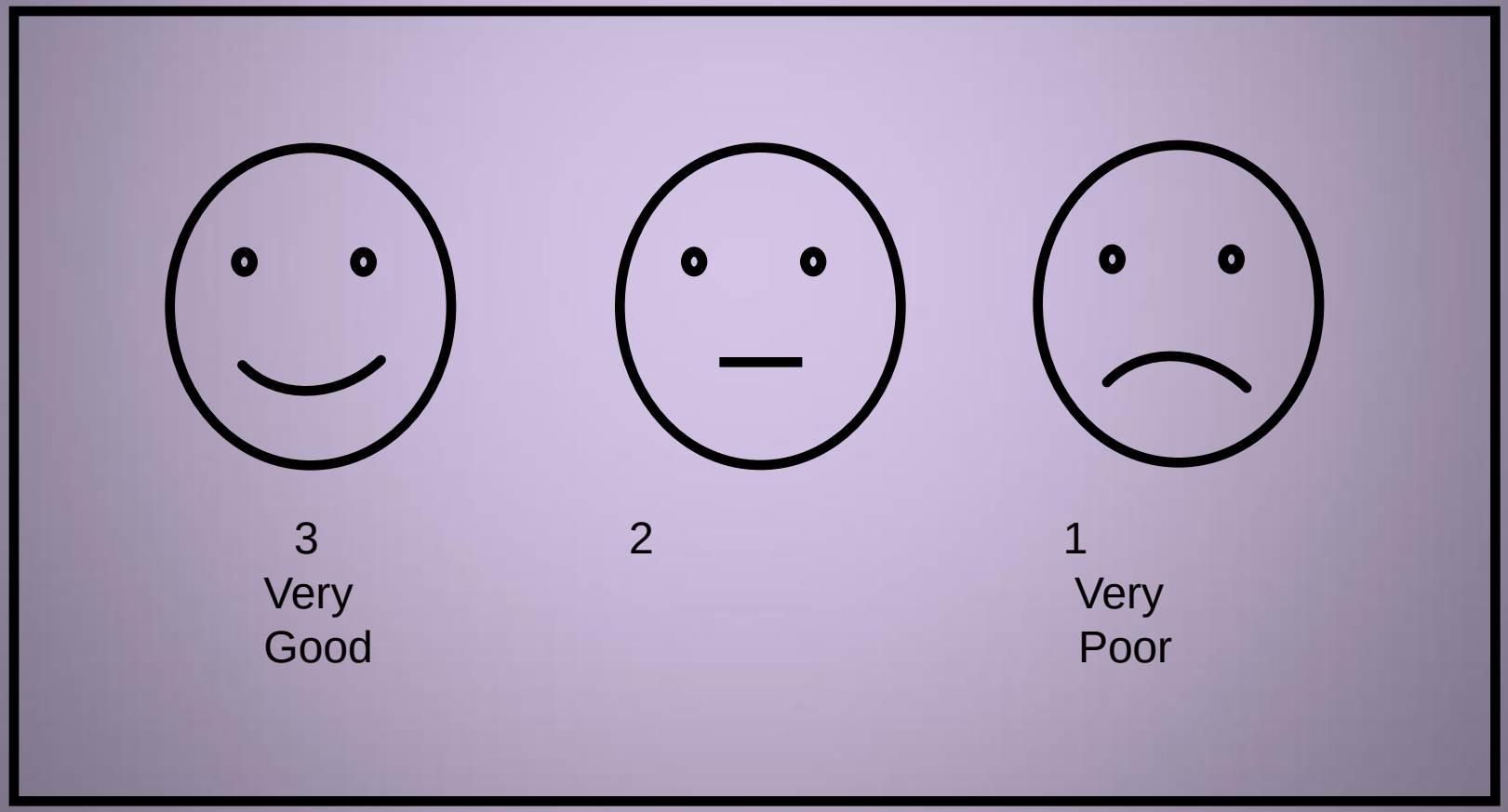
Type D



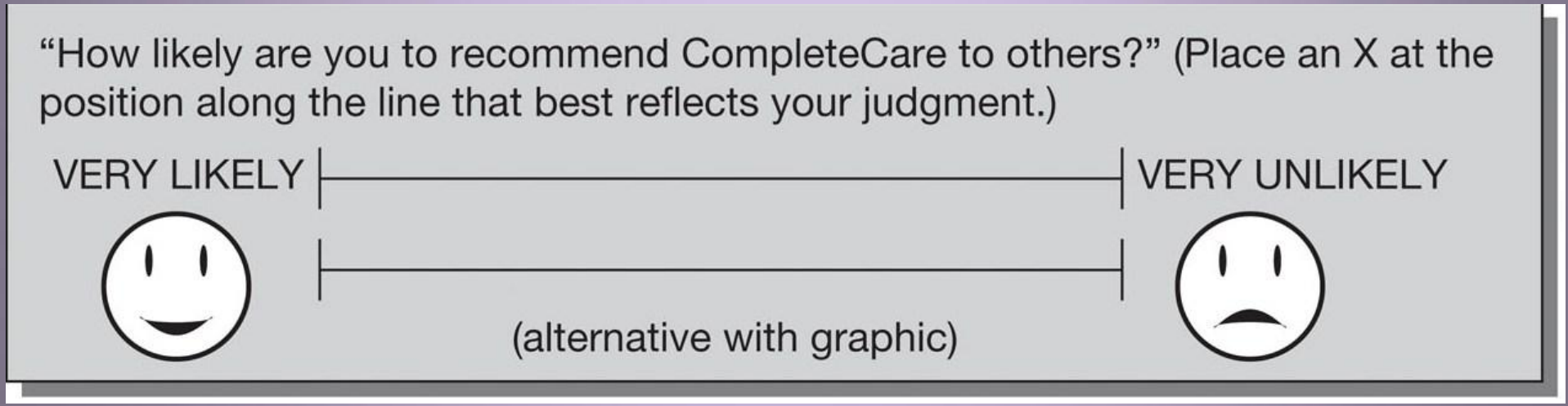
Graphic Rating Scales

- A graphic rating scale presents respondents with a graphic continuum.

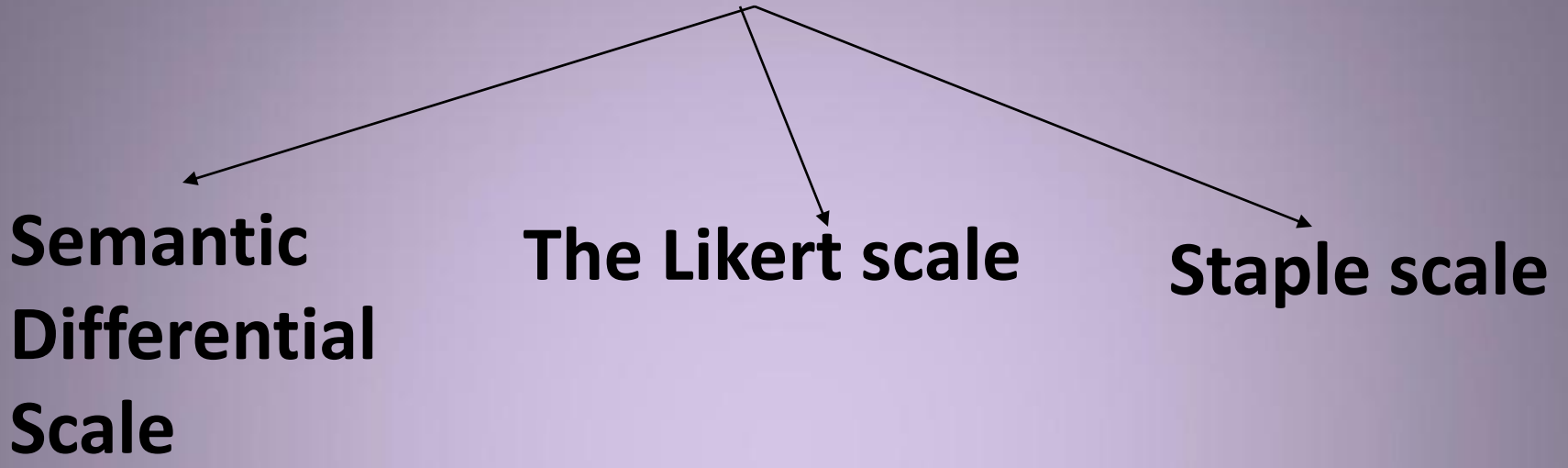
Graphic Rating Scale Stressing Pictorial Visual Communications



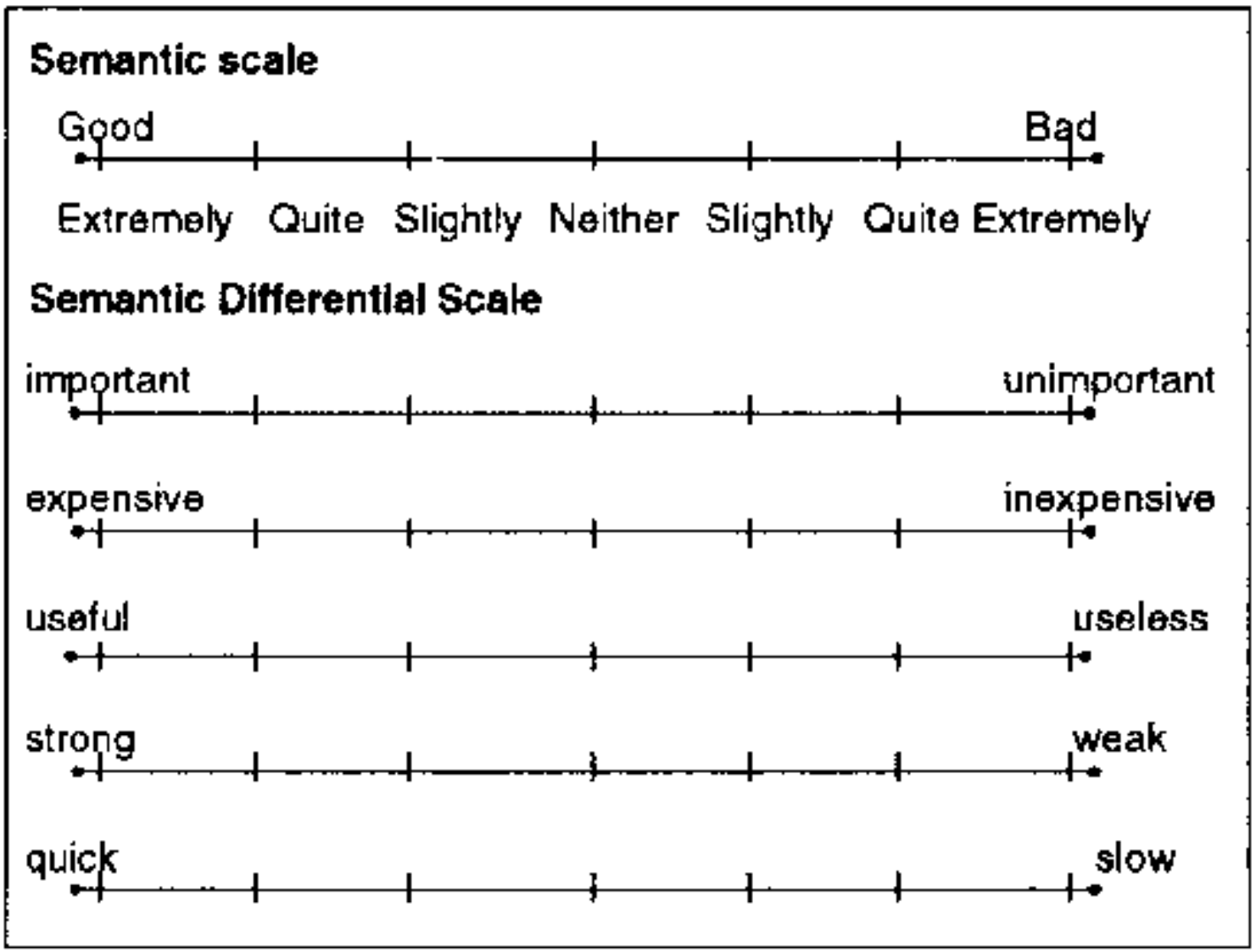
Graphic Rating Scales



Itemized Rating Scales



Non-Comparative Scales



Non-Comparative Scales

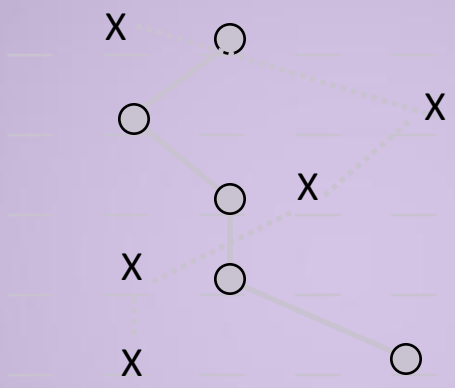
Semantic Differential Scale

Here are a number of statements that could be used to describe K-Mart. For each statement tick (X) the box that best describes your feelings about K-Mart.

Modern Store	☐	☐	☐	☐	☐	☐	☐	**Old- fashioned store**
Low prices	☐	☐	☐	☐	☐	☐	☐	**High prices**
Unfriendly staff	☐	☐	☐	☐	☐	☐	☐	**Friendly staff**
Narrow product range	☐	☐	☐	☐	☐	☐	☐	**Wide product range**
Sophisticated customers	☐	☐	☐	☐	☐	☐	☐	**Unsophisticated customers**

Semantic Differential Scale - Snake Diagram

Modern Store		Old- fashioned store
Low prices		High prices
Friendly staff		Unfriendly staff
Wide product range		Narrow product range
Sophisticated customers		Unsophisticated customers

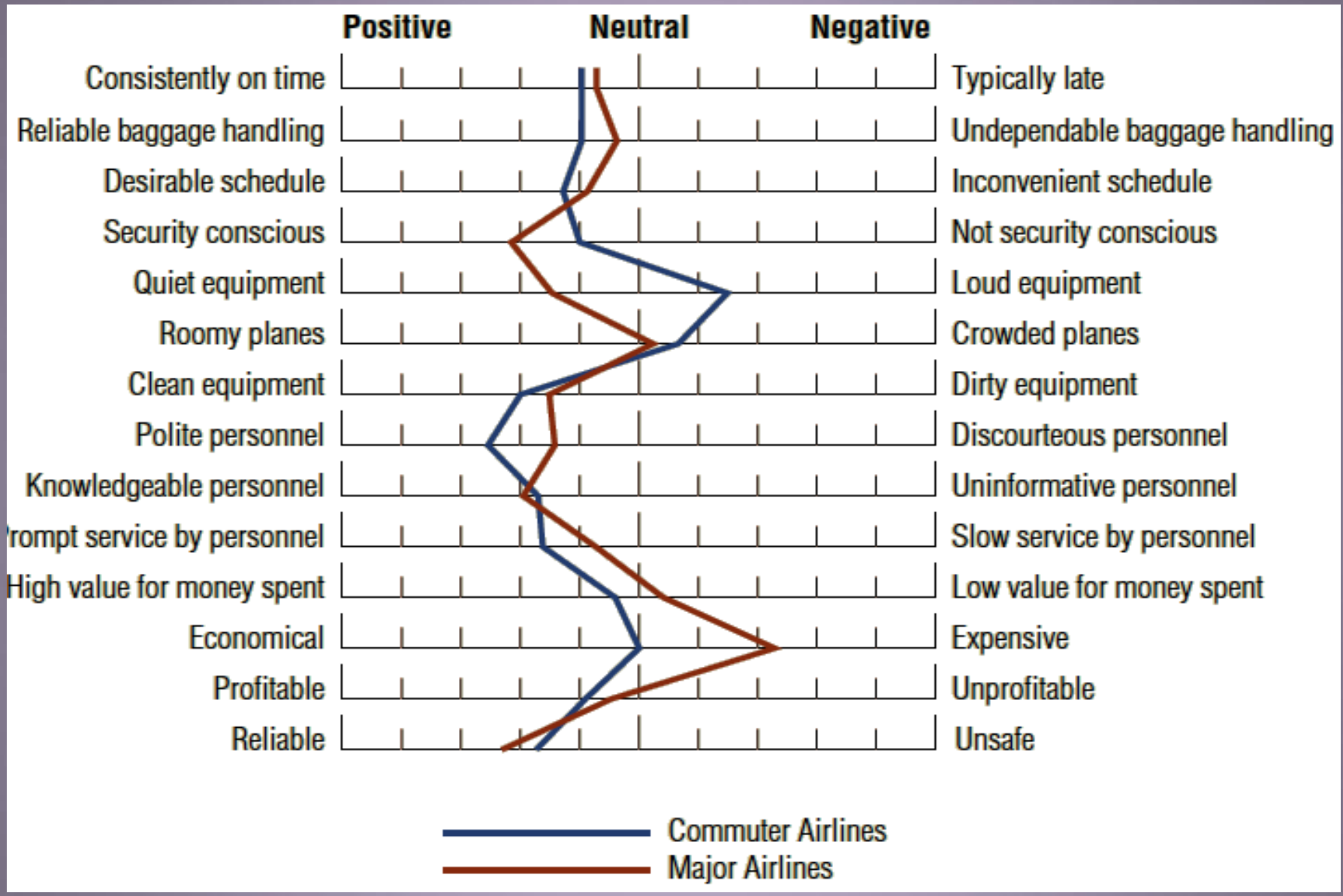


Key :

○ Sears

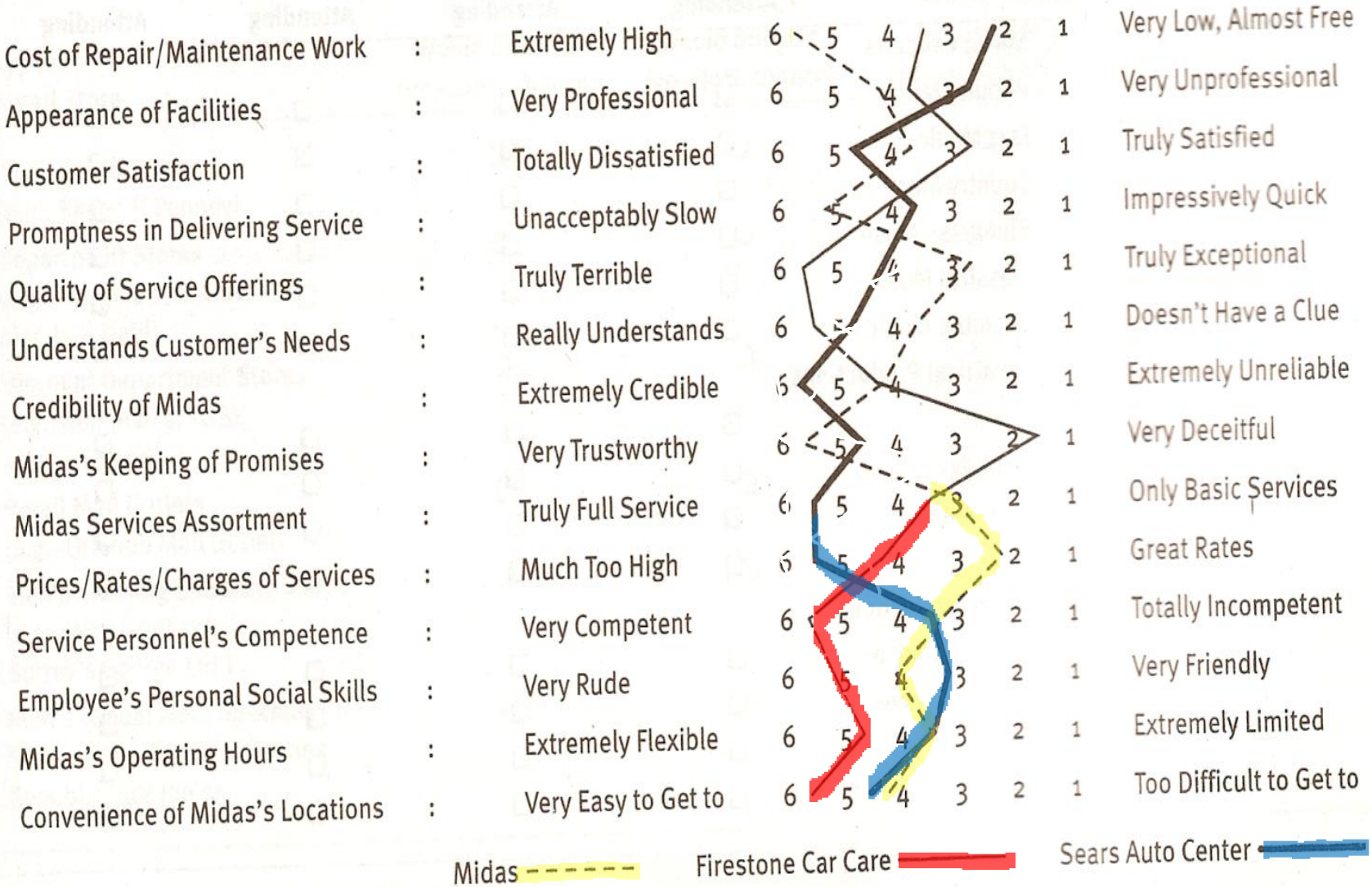
X K-Mart

Image Profile of Commuter Airlines versus Major Airlines



Source: J. Richard Jones and Sheila I. Cocke, "A Performance Evaluation of Commuter Airlines: The Passengers' View," *Proceedings:* Transportation Research Forum 22 (1981), p. 524. Reprinted with permission.

BRAND PROFILING – CAR CARE



Itemised Rating Scales

Semantic differential

Old Fashioned	**X** **1**	**2**	**3**	**4**	**5**	Modern
Cheap	**1**	**2**	**3**	**X** **4**	**5**	Expensive
Friendly service	**1**	**2**	**X** **3**	**4**	**5**	Unfriendly service

Itemised Rating Scales

Likert scale

	Strongly agree	disagree	Neither agree nor disagree	agree	Strongly agree
Market research is the most interesting subject known to man	1	2	3	4	5

Itemised Rating Scales

The Likert scale

For each of the statements below, please indicate the extent of your agreement or disagreement by placing a tick in the appropriate column	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
1. Criminals convicted of murder should be hanged					
2. Trial by jury should be abolished					

What is Likert scale?

- It is a psychometric scale commonly involved in research that employs questionnaires.
- It is the most widely used approach to scaling responses in survey research.
- Likert scales are a non-comparative scaling technique and are one-dimensional in nature.
- When responding to a Likert questionnaire item respondents specify their level of agreement or disagreement on a symmetric agree-disagree scale for a series of statements.
- Thus, the range captures the intensity of their feelings for a given item, while the results of analysis of multiple items reveals a pattern that has scaled properties of the kind Likert identities

Five – point Likert item

Likert Scale

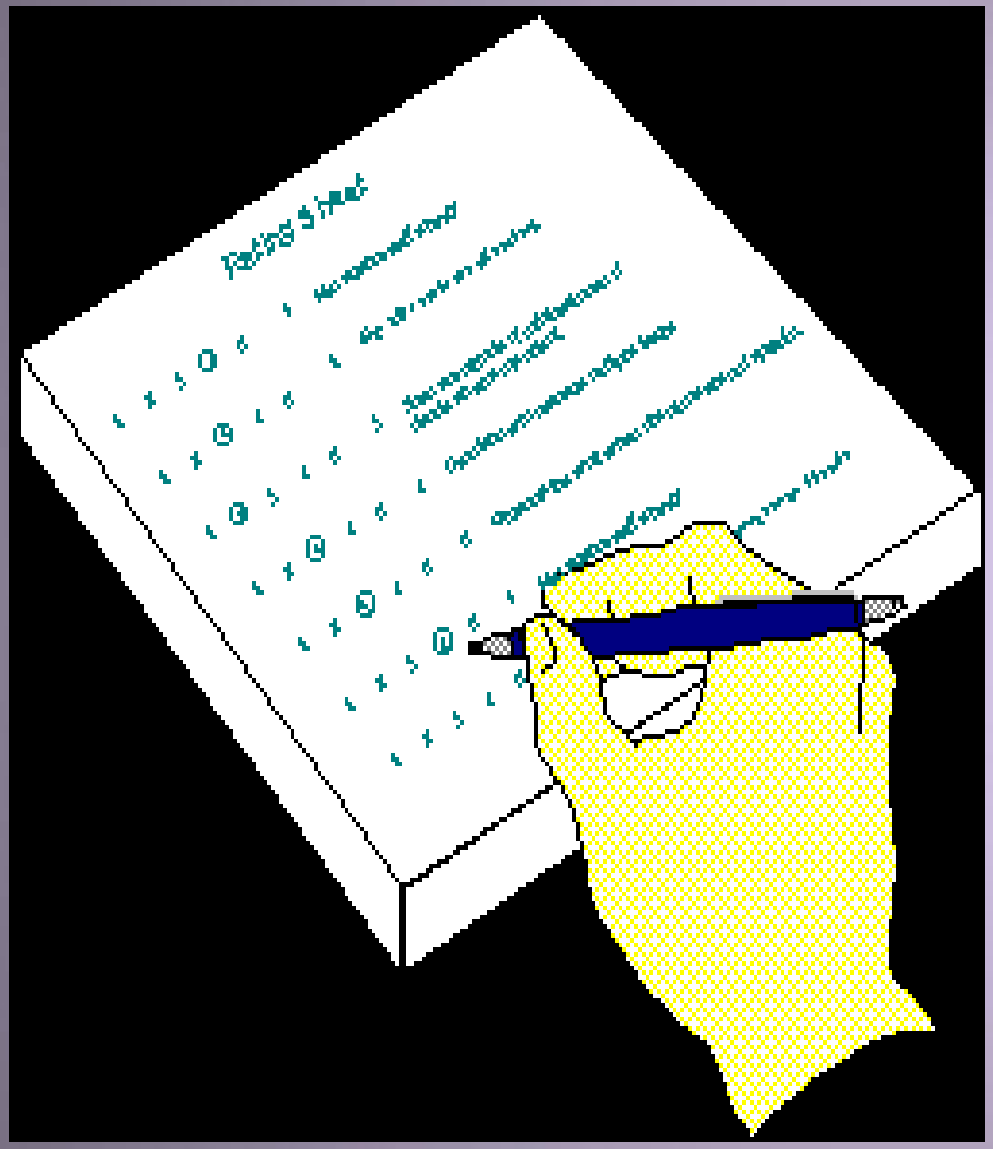
Likert item

Difference

- **Likert item is considered symmetric or balanced because there are equal amounts of positive and negative positions.**

- **Often five ordered response levels are used, although many psychometricians advocate using seven or nine level, a recent empirical study found that a 5 or 7 point scale.**

The format of a typical five-level Likert item



1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

Example:

Q.18.Please measure the following affirmative perceptions about your library OPAC and Web OPAC use .

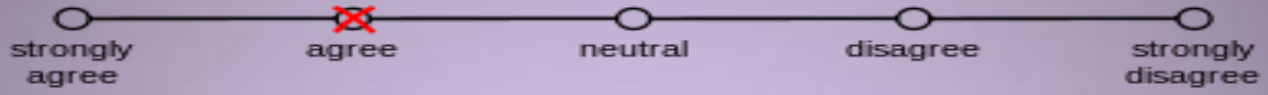
Codes: 1-strongly disagree, 2- disagree, 3- neutral, 4-agree, 5- strongly agree						
S/N	Affirmative perception statements	**Codes**				
(i)	OPAC / Web OPAC was Easier to use than I expected	1	2	3	4	5
(ii)	It was Fun to use	1	2	3	4	5
(iii)	It was Easy to use	1	2	3	4	5
(iv)	It helped me in finding the documents faster	1	2	3	4	5
(v)	It is Very difficult to use	1	2	3	4	5
(vi)	It is Very confusing to use	1	2	3	4	5
(vii)	I found more items than expected	1	2	3	4	5
(viii)	I am comfortable with simple search	1	2	3	4	5
(ix)	I am comfortable complex/Advance search	1	2	3	4	5
(x)	I am comfortable quick search	1	2	3	4	5
(xi)	I am comfortable when using OPAC/Web OPAC	1	2	3	4	5

Q.19. Please mark your appreciation towards the use of your library OPAC/Web OPAC.

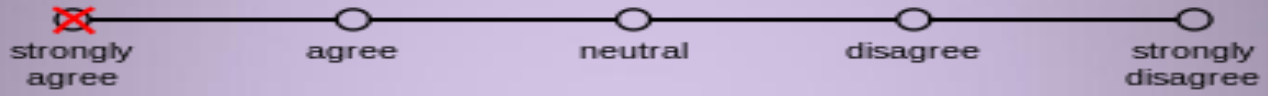
Codes: 1-strongly disagree, 2- disagree, 3- neutral, 4-agree, 5- strongly agree						
S/N	Parameters	Codes				
(i)	I access OPAC/Web OPAC stand alone system	1	2	3	4	5
(ii)	I access OPAC/Web OPAC library premises	1	2	3	4	5
(iii)	It is easy to be familiar with this OPAC/Web OPAC	1	2	3	4	5
(iv)	The OPAC/Web OPAC should have more flexible interfaces	1	2	3	4	5
(v)	Library searching will be easier and faster with the Web OPAC	1	2	3	4	5
(vi)	It is easy to read information provided in the Web OPAC	1	2	3	4	5
(vii)	A OPAC/Web OPAC search by author is easy	1	2	3	4	5
(viii)	A OPAC/ Web OPAC search by call number is easy	1	2	3	4	5
(ix)	OPAC/ Web OPAC scanning through a long display (forward or backward) is easy	1	2	3	4	5
(x)	OPAC/ Web OPAC reducing the result when too much is retrieved is easy	1	2	3	4	5

Example Likert Scale

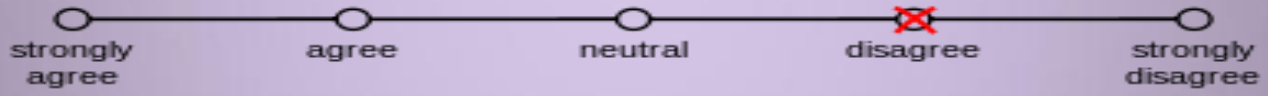
1. Wikipedia has a user friendly interface.



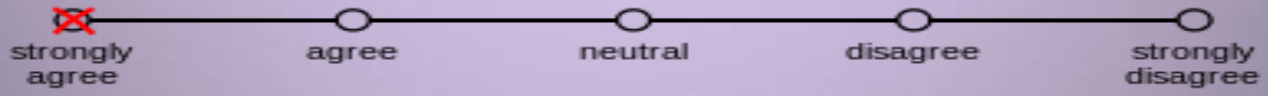
2. Wikipedia is usually my first resource for research.



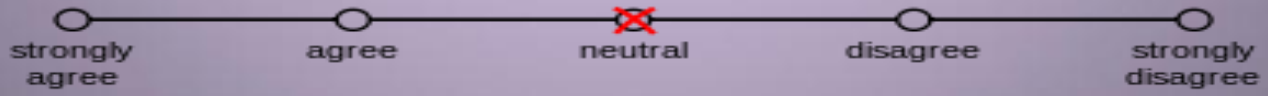
3. Wikipedia pages generally have good images.



4. Wikipedia allows users to upload pictures easily.



5. Wikipedia has a pleasing color scheme.



The format of a typical Seven-level Likert item

1. It is the duty of doctors to keep people alive for as long as possible.

- ☐ Strongly Agree
- ☐ Agree
- ☐ Agree somewhat
- ☐ Undecided
- ☐ Disagree somewhat
- ☐ Disagree
- ☐ Strongly disagree

- Formal assessments of risk should be made accessible to all members of the interdisciplinary team

1	2	3	4	5	6	7	8	9
Disagree								Agree

- Risk assessment tools should be used when assessing the risk of an individual for developing pressure ulcer(s)

1	2	3	4	5	6	7	8	9
Disagree								Agree

- Clinical judgement can be as accurate at predicting risk as assessment tool scores

1	2	3	4	5	6	7	8	9
Disagree								Agree

- Risk assessment tools should only be used in conjunction with clinical judgement when assessing the risk of an individual for developing pressure ulcer(s)

1	2	3	4	5	6	7	8	9
Disagree								Agree

In a 20 year career, it is more difficult for a seaman recruit to advance to master chief petty officer than it is for an ensign to advance to commander.

Analysis Methods

Depending on how the Likert scale questions are treated a number of different analysis methods can be applied

1. Analysis methods used for individual questions (ordinal data)
   - Bar charts and dot plots
     - Not histograms (data is not continuous)
   - Central tendency summarised by median and mode
     - Not mean
   - Variability summarised by range and interquartile range
     - Not standard deviation
   - Analysed using non-parametric tests
     (difference between the medians of comparable groups)
     - Mann- whitney U test
     - Wilcoxon signed –rank test
     - Kruskal – wallis test

2. When multiple Likert question responses are summed together (interval data)

- All questions must use the same Likert scale
- Must be a defendable approximation to an interval scale (i.e. coding indicates magnitude of difference between items but there is no absolute zero point)
- All items measure are single latent variable (i.e. a variable that is not directly observed, but rather inferred from other variables that are observed and directly measured)
- Analyzed using parametric tests
    - Analysis of variance (ANOVA)

3. Analysis methods used when reduced to nominal level of agree vs. disagree

- Chi –square test
- Cochran Q test
- McNemar test

Advantages

- Item analysis increases the degree of homogeneity or internal consistency in the set of statements.
- Subjects generally find it easy to respond because they have a wide range of answers(usually five) to choose from instead of only two alternative responses, i.e., agree or disagree.
- No outside group of judges is involved in selecting statements and giving values to them.

Limitations

- Ties in ranks occur quite frequently.
- The response pattern of an individual is not revealed.
- A respondent is required to answer all questions on the scale.
- A problem of interpretation arises with this type of scale.
- In this scale all statements of a universe are deemed to be of equal attitude value.

Conclusion

A summated rating scale is a set of attitude statements all of which subjects respond with degrees of agreement or disagreement carrying different scores. These scores are summed or summed and averaged to yield an individual's attitude score. The objective is to avoid the use of only a single statement to represent a concept and instead to use several statement as indicators, all representing different facets of the concept to obtain a more well rounded perspective.

AGREEMENT

•Strongly Agree **•Agree** **•Undecided** **•Disagree** **•Strongly Disagree**	•Agree Strongly •Agree Moderately •Agree Slightly •Disagree Slightly •Disagree Moderately •Disagree Strongly	•Agree •Disagree	•Agree •Undecided •Disagree
•Agree Very Strongly •Agree Strongly •Agree •Disagree •Disagree Strongly •Disagree Very Strongly	•Yes •No	•Completely Agree •Mostly Agree •Slightly Agree •Slightly Disagree •Mostly Disagree •Completely Disagree	•Disagree Strongly •Disagree •Tend to Disagree •Tend to Agree •Agree •Agree Strongly

FREQUENCY

•Very Frequently •Frequently •Occasionally •Rarely •Very Rarely •Never	•Always •Very Frequently •Occasionally •Rarely •Very Rarely •Never	•Always •Usually •About Half the Time •Seldom •Never	•Almost Always •To a Considerable Degree •Occasionally •Seldom
•A Great Deal •Much •Somewhat •Little •Never	•Often •Sometimes •Seldom •Never	•Always •Very Often •Sometimes •Rarely •Never	

IMPORTANCE

•Very Important •Important •Moderately Important •Of Little Importance •Unimportant	•Very Important •Moderately Important •Unimportant

QUALITY

•Very Good •Good •Barely Acceptable •Poor •Very Poor	•Extremely Poor •Below Average •Average •Above Average •Excellent	•Good •Fair •Poor

LIKELIHOOD

•**Like Me** •**Unlike Me**	•To a Great Extent •Somewhat •Very Little •Not at All	•True •False
•Definitely •Very Probably •Probably •Possibly •Probably Not •Very Probably Not	•**Almost Always True** •**Usually True** •**Often True** •**Occasionally True** •**Sometimes But Infrequently True** •**Usually Not True** •**Almost Never True**	•True of Myself •Mostly True of Myself •About Halfway True of Myself •Slightly True Of Myself •Not at All True of Myself

Itemised Rating Scales

Staple scale

+5	+5
+4	+4
+3	+3
+2	+2
+1	+1
High quality	Poor service
-1	-1
-2	-2
-3	-3
-4	-4
-5	-5

A Stapel Scale for Measuring a Store's Image

Select a plus number for words that you think describe the store accurately. The more accurately you think the work describes the store, the larger the plus number you should choose. Select a minus number for words you think do not describe the store accurately. The less accurately you think the word describes the store, the larger the minus number you should choose, therefore, you can select any number from +3 for words that you think are very accurate all the way to -3 for words that you think are very inaccurate.

+3
+2
+1
Wide Selection
-1
-2
-3

Staple Scale

The following questions concern your ratings of several suppliers that provide products for use in your store.

XYZ

Poor Product Selection	-5	-4	-3	-2	-1	1	2	3	4	5
Costly Products	-5	-4	-3	-2	-1	1	2	3	4	5
Fast Service	-5	-4	-3	-2	-1	1	2	3	4	5
High Quality Products	-5	-4	-3	-2	-1	1	2	3	4	5
Innovative	-5	-4	-3	-2	-1	1	2	3	4	5

Table 9.3 Stapel Scale

Abundance of Exhibits	Complex User Registration	Low Commission	Good Response to Complaints	Poor Protection of Personal Information	Confusing Support System
+5	+5	+5	+5	+5	+5
+4	+4	+4	+4	+4	+4
+3	+3	+3	+3	+3	+3
+2	+2	+2	+2	+2	+2
+1	+1	+1	+1	+1	+1
-1	-1	-1	-1	-1	-1
-2	-2	-2	-2	-2	-2
-3	-3	-3	-3	-3	-3
-4	-4	-4	-4	-4	-4
-5	-5	-5	-5	-5	-5

Monadic Rating Scale

A **Monadic Rating Scale** asks about a single concept

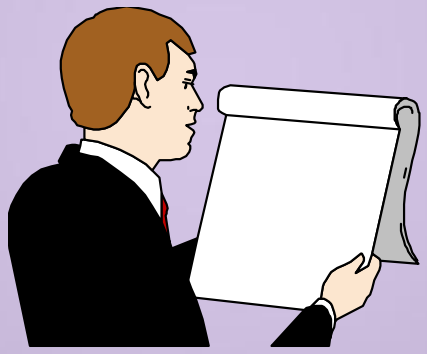
Now that you've had your automobile for about 1 year, please tell us how satisfied you are with its engine power and pickup.

Completely Satisfied	**Very Satisfied**	**Fairly Well Satisfied**	**Somewhat Dissatisfied**	**Very Dissatisfied**

Some Basic Considerations When Selecting a Scale

Selecting a Rating, Ranking, Sorting, or Purchase Intent Scale

Number of Categories



Odd or Even Number of Scale Categories

Forced Versus Non-forced Choice

Balanced Versus Non-balanced Alternatives

Odd versus even

if neutral responses likely, use odd number

Odd		Even	
Strongly Agree	_____	Strongly Agree	_____
Agree	_____	Agree	_____
Neutral	_____	Disagree	_____
Disagree	_____	Strongly disagree	___
Strongly disagree	_____		

Balanced vs. Unbalanced

Balanced		Unbalanced	
Very good	_____	Excellent	_____
Good	_____	Very Good	_____
Fair	_____	Good	_____
Poor	_____	Fair	_____
Very Poor	_____	Poor	_____

Balanced and Unbalanced Scales

Balanced Scale	Unbalanced Scale
JOVAN MUSK FOR MEN IS	*JOVAN MUSK FOR MEN IS*
Extremely good	Extremely good
Very good	Very good
Good	Somewhat Good
Bad	Good
Very bad	Bad
Extremely bad	Very bad

Forced vs. Unforced

Forced

Extremely Reliable	____
Very Reliable	____
Somewhat Reliable	____
Somewhat Unreliable	____
Very Unreliable	____
Extremely Unreliable	____

Unforced

Extremely Reliable	____
Very Reliable	____
Somewhat Reliable	____
Somewhat Unreliable	____
Very Unreliable	____
Extremely Unreliable	____
Don’t know	____

Labeled vs. End Anchored

Labeled

Excellent _____

Very Good _____

Fair _____

Poor _____

Very Poor _____

End Anchored

Excellent _____

_____

_____

_____

Poor _____

Intervals May Not Reflect the Semantic Meaning of the Adjectives

Excellent _____

Labeled

Excellent _____

Very Good _____

Fair _____

Poor _____

Very Poor _____

Intervals Are Not Equal

Very Good _____

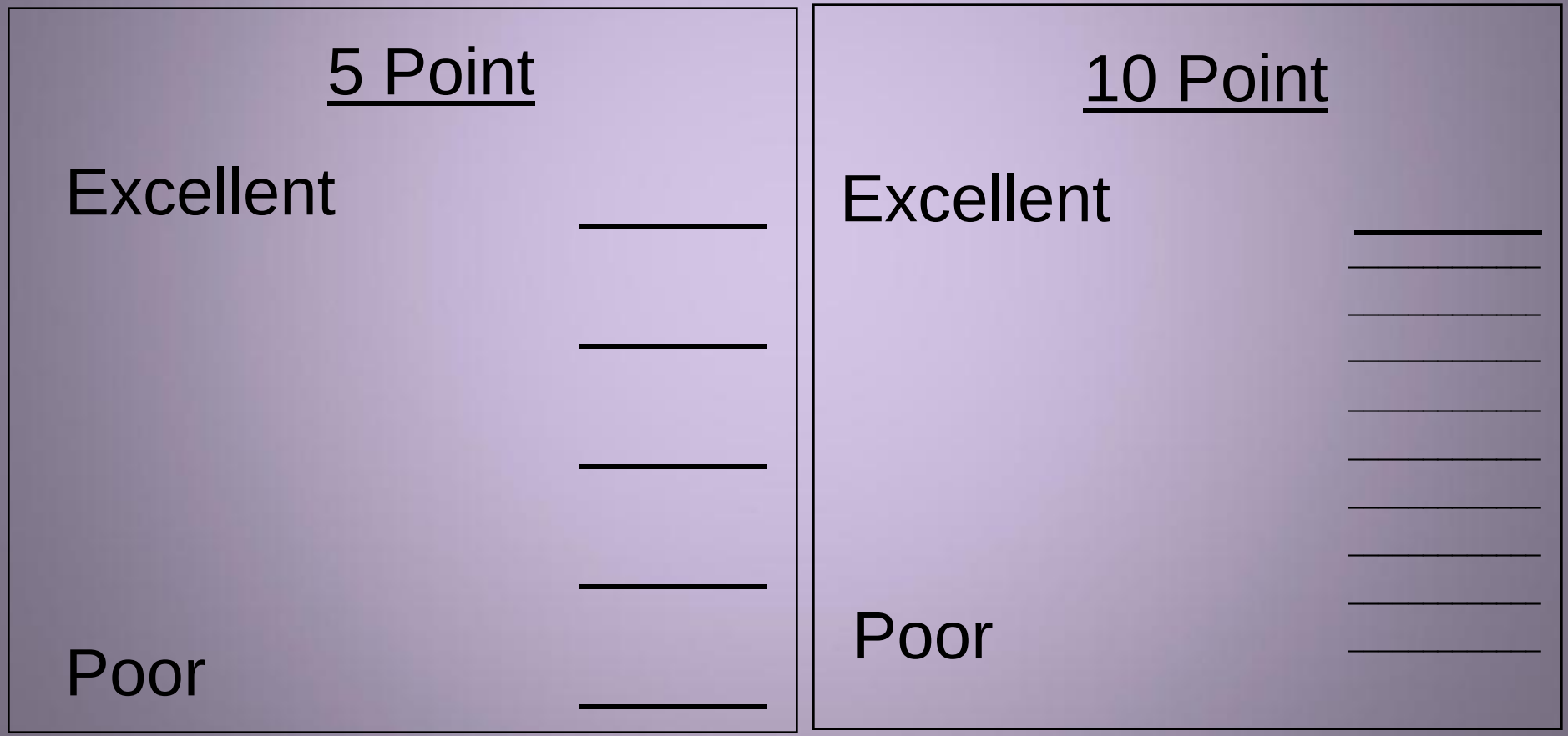
Fair _____

Poor _____

Intervals Are Not Equal

Very Poor _____

Number of Scale Points



Choosing the Appropriate Scale

Attitude component	Itemized category	Rank order	Constant sum	Likert	Semantic differential
Knowledge					
Awareness	A				
Attribute beliefs	A	B	B	B	A
Attribute importance	A	B	A	B	
Affect or Liking					
Overall preferences	A	B	A	B	B
Specific attributes	A	B	B	B	A
Action					
intentions	A	B	A	B	

A = Very appropriate, B = Sometimes appropriate

Thurston Scales

- Thurston Scales
  - Items are formed
  - Panel of experts assigns values from 1 to 11 to each item
  - Mean or median scores are calculated for each item
  - Select statements evenly spread across the scale

Thurston Scales

- Example:

  Please check the item that best describes your level of willingness to try new tasks

  – I seldom feel willing to take on new tasks (1.7)
  – I will occasionally try new tasks (3.6)
  – I look forward to new tasks (6.9)
  – I am excited to try new tasks (9.8)

Guttman Scales

- Also known as Scalograms
- Both the respondents and items are ranked
- Cutting points are determined (Good enough-Edwards technique)
- Coefficient of Reproducibility (CR_{eg}) - a measure of goodness of fit between the observed and predicted ideal response patterns
- Keep items with CR_{eg} of 0.90 or higher

Guttman Scales

- also known as **cumulative scaling** or **scalogram analysis**.
- Goal: a set of items or statements such that a respondent who agrees with any specific question in the list will also agree with all previous statements.

Guttman

- Same as before…generate the items
- Judges rate the statements or items in terms of how favorable they are to the concept of immigration (Yes=favorable; No=unfavorable)
- The key to Guttman scaling is in the analysis.

Guttman Scaling

- construct a matrix that shows the responses of all the judges on all of the items
- sort this matrix so that respondents who agree with more statements are listed at the top and those agreeing with fewer are at the bottom.
- The resulting scale should be nearly cumulative

Guttman Matrix

Judge	Item 7	Item 4	Item 9	Item 2	Item 3	Item k
5	Y	Y	Y	Y	Y	…
8	Y	Y	Y	**N**	**Y**	N
2	Y	Y	Y	Y	N	N
6	Y	Y	Y	Y	N	N
17	Y	Y	Y	N	N	N
32	Y	Y	N	N	N	N
22	Y	Y	N	**Y**	N	N
15	Y	Y	N	N	N	N
4	Y	N	N	**Y**	N	N
9	N	N	N	N	N	N

Exceptions are circled

Guttman Scaling

- Can do this for a small number of items
- if there are many items, then a **scalogram analysis** is used to do the same thing
  - determines the subsets of items from our pool that best approximate the cumulative property
  - many analyses used to determine how well the items approximate a true cummulative scale

Guttman Scale

- **INSTRUCTIONS:** Place a check next to each statement you agree with.
- _____ I would permit a child of mine to marry an immigrant.
- _____ I believe that this country should allow more immigrants in.
- _____ I would be comfortable if a new immigrant moved next door to me.
- _____ I would be comfortable with new immigrants moving into my community.
- _____ It would be fine with me if new immigrants moved onto my block.
- _____ I would be comfortable if my child dated a new immigrant.

Each item has a scale value associated with it (obtained from the scalogram analysis).
A respondent's scale score is the sum the scale values of every item they agree with.

Mokken Scaling

- Just like Guttman scales
- Individual's scale score is the rank of the highest item endorsed or the total number of items endorsed
- Key difference
  - Mokken scales are probabilistic
  - A respondent answering an item positively will have a large probability of answering a less difficult item in a positive way as well

Multidimensional Scaling

- Many interesting concepts are not unidimensional
  - (e.g., attitude toward immigration)
- MDS maps the distances between points in a high dimensional space into a lower dimensional space without too much loss of information.
- Almost always 2-D

Multidimensional Scaling

- The “points” that are represented in multidimensional space can be just about anything.

- These objects might be people, in which case MDS can identify clusters of people who are “close” versus “distant” in some real or psychological sense.

Multidimensional Scaling

- The objects might be physical objects (e.g., cities, cars, books) or psychological or cultural objects (e.g., personality traits, religions, political parties).

- The only requirement is that some basis exist for rating or ranking the objects in terms of similarity

Multidimensional Scaling

- As long as the “distance” between the objects can be assessed in some fashion, MDS can be used to find the lowest dimensional space that still adequately captures the distances between objects.
- Once the number of dimensions is identified, must identify the meaning of the dimensions.

Multidimensional Scaling

- Basic data representation in MDS is a *dissimilarity matrix* that shows the distance between every possible pair of objects
- The goal of MDS is to faithfully represent these distances with the lowest possible dimensional space
- We seek a reference system that can capture the basic information in a data set—in this case a data set characterized by distances between objects.

Multidimensional Scaling

- This table lists the distances between European cities. A multidimensional scaling of these data should be able to recover the two dimensions (North-South x East-West) that we know must underlie the spatial relations among the cities.

	Athens	Berlin	Dublin	London	Madrid	Paris	Rome	Warsaw
Athens	0	1119	1777	1486	1475	1303	646	1013
Berlin	1119	0	817	577	1159	545	736	327
Dublin	1777	817	0	291	906	489	1182	1135
London	1486	577	291	0	783	213	897	904
Madrid	1475	1159	906	783	0	652	856	1483
Paris	1303	545	489	213	652	0	694	859
Rome	646	736	1182	897	856	694	0	839
Warsaw	1013	327	1135	904	1483	859	839	0

Multidimensional Scaling

- In the cities data, the meaning is quite clear.
- The dimensions refer to the North-South x East-West surface area across which the cities are dispersed.
- Expect MDS to faithfully recreate the map relations among the cities.

Multidimensional Scaling

- Common to evaluate the fit of a model using a "stress" statistic

Stress	***Fit***
.20	Poor
.10	Fair
.05	Good
.02	Excellent

Multidimensional Scaling

- Stress Statistic

given: - a set of n objects
- the dissimilarities δ_{ij} between them

find: points on the plane whose distances d_{ij} are as close as possible to the δ_{ij}

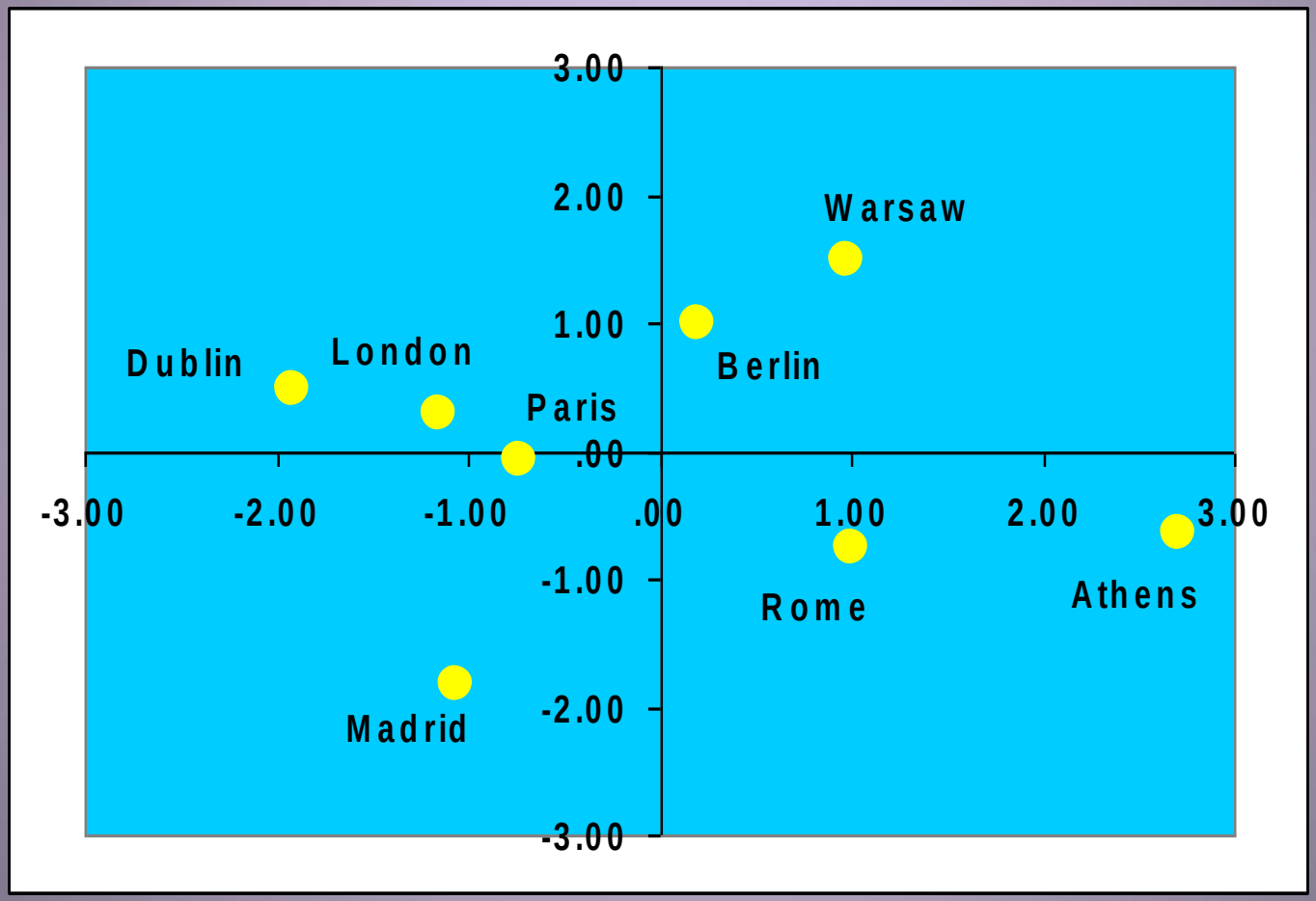
minimize: $$\text{STRESS} = \left[\frac{\sum_{i,j} \left(d_{ij} - \delta_{ij} \right)^2}{\sum_{i,j} \delta_{ij}^{\,2}} \right]^{1/2}$$ [Kruskal 1964]

- missing values: only ~20 distances/object are needed

Multidimensional Scaling

- The stress for the one-dimensional model of the cities data is .31, clearly a poor fit
- The analyses proceeds by trying higher dimensional models, seeking the minimum number of dimensions necessary to get an acceptable fit
- A 2-D model fits very well (stress value =.009) indicating an exceptional fit. This is no great surprise for these data.

Multidimensional Scaling



Multidimensional Scaling

- Rarely do social sciences have metric data on which to compare the objects. More typically, the “distances” are on psychological dimensions for which ordinal relations are the best scaling possible
- Use similarity ratings instead
  - Similarity is a slippery slope
  - Can be very time consuming because they require the rating of each pair of objects

Characteristics of Good Measurement Scales

1. Reliability

- **The degree to which a measure accurately captures an individual's true outcome without error; *Accuracy***
- **synonymous with repetitive consistency**

2. Validity

- **The degree to which a measure faithfully represents the underlying concept; *Fidelity***

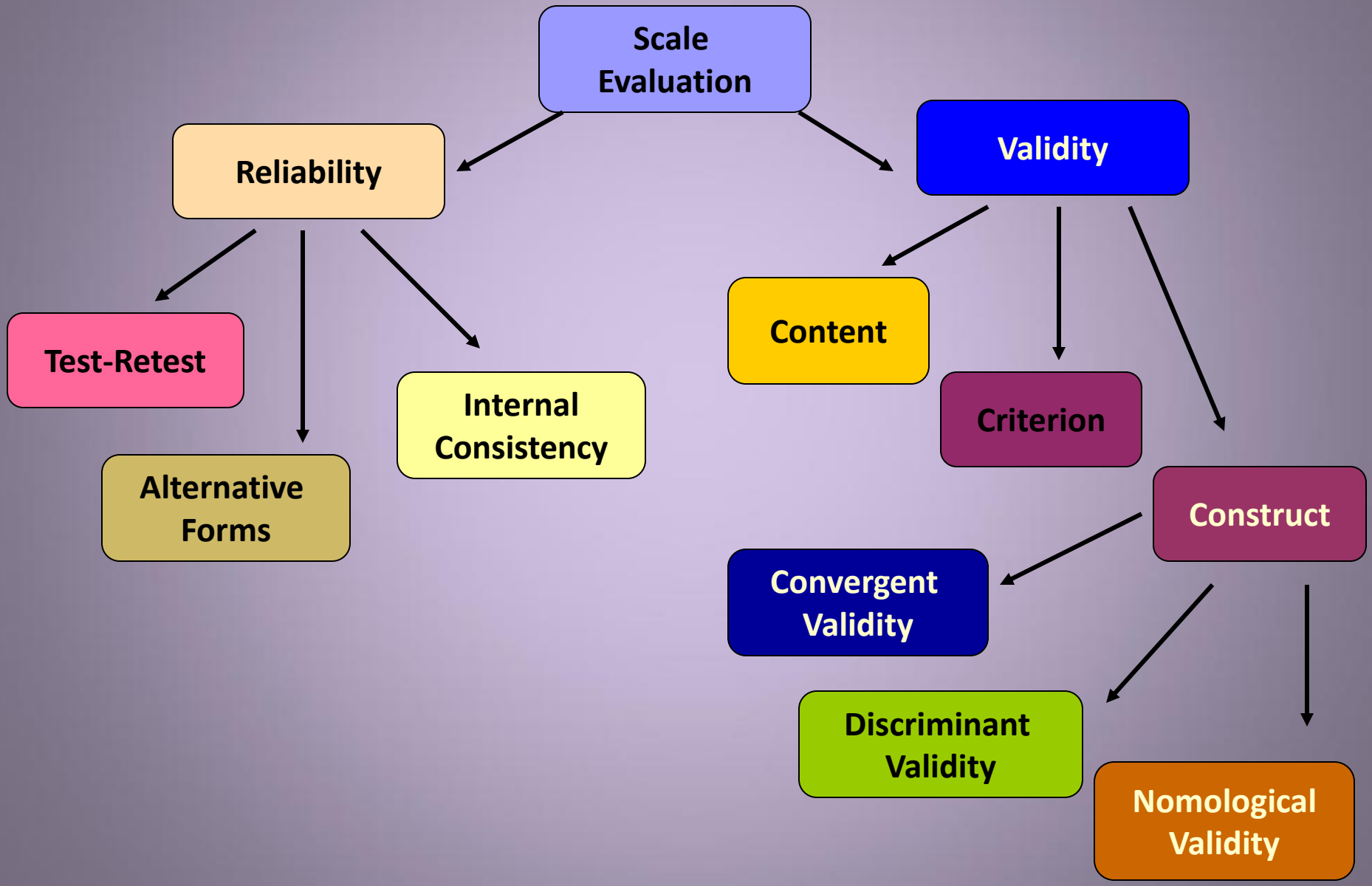
3. Sensitivity

- **The ability to discriminate meaningful differences between attitudes. The more categories the more sensitive (but less reliable)**

4. Generalizability

- **How easy is scale to administer and interpret**

Figure 10.7 Scale Evaluation



Reliability and Validity

- Reliability: Refers to the degree of variable error represented in a measurement
- Validity: Refers to the extent to which the measurement is free from systematic error Perfect validity suggests M=C; no E.

Key Idea

- To develop reliable and valid measures which we can subsequently (and appropriately) use in statistical analyses, we must understand:
  - Properties of scales
  - How to design good questions that do not lead to biased or inconsistent responses

Different Types of Reliability

- **Internal Reliability**
  - Extent to which items on a scale "hang together" or are correlated with one another
    - Cronbach's alpha
    - Split-half reliability (split measure into halves, correlate)

- **Test-Retest Reliability**
  - Extent to which scores are stable over time
  - Have people completed questionnaire twice and correlate scores

Approaches to Assess Reliability

- Test-Retest: Repeated measurement of the same person or group using the same scale under the similar conditions are taken
- Split Sample: The number of items is randomly divided into two parts and a correlation coefficient between the two is obtained. A high correlation indicates the internal consistency of the construct which leads to greater reliability
- Alternative Forms
- Internal Comparison (Cronbach's Coefficient Alpha)
- Inter-Scorer Reliability

Reliability

- **Reliability** can be defined as the extent to which measures are free from random error.
- In **test-retest reliability**, respondents are administered identical sets of scale items at two different times and the degree of similarity between the two measurements is determined.
- In **alternative-forms reliability**, two equivalent forms of the scale are constructed and the same respondents are measured at two different times, with a different form being used each time.

Reliability

- **Internal consistency reliability** determines the extent to which different parts of a summated scale are consistent in what they indicate about the characteristic being measured.

- In **split-half reliability**, the items on the scale are divided into two halves and the resulting half scores are correlated.
- Cronbach's Alpha1 (*α) is a measure of internal consistency that is popular* in the field of psychometrics.
- Let *k be the number of items (or questions),* s_i^2 *is* the variance associated with item *i,* and s_{T}^2 *is* the variance associated with the total (or sum) of all *k item scores. Cronbach's* alpha is mathematically defined as follows:

- $$\alpha = \frac{k}{k-1}\left(1 - \frac{1}{s_{\mathrm{T}}^2}\sum_{i=1}^{k} s_i^2\right)$$

- The **coefficient alpha**, or Cronbach's alpha, is the average of all possible split-half coefficients resulting from different ways of splitting the scale items. This coefficient varies from 0 to 1, and a value of 0.6 or less generally indicates unsatisfactory internal consistency reliability. Squaring this correlation and subtracting from 1.00 produces the index of measurement error. As the estimate of reliability increases, the fraction of a test score that is attributable to error will decrease.

Computing Cronbach's Alpha

Table 11.1: Leadership Data from 15 Respondents

	Questions/Items						
Respondent	1A	1B	1C	1D	1E	1F	Total
1	5	5	3	3	2	2	20
2	4	4	3	5	4	4	24
3	4	5	4	4	4	4	25
4	5	3	4	4	2	2	20
5	4	3	5	4	3	3	22
6	5	2	1	1	1	1	11
7	5	5	4	4	4	4	26
8	4	4	3	2	1	1	15
9	5	5	1	1	1	1	14
10	5	4	1	2	2	1	15
11	4	4	3	4	3	2	20
12	3	2	3	4	3	4	19
13	5	5	5	1	4	1	21
14	5	5	3	2	4	4	23
15	5	4	3	2	4	2	20

Computing Cronbach's Alpha

- To illustrate the calculation of Cronbach's alpha, let us consider the data in Table 11.1. As previously mentioned, the number of items being studied is *k = 6. The 6 item-level* variances are given by $s_1^2 = 0.4095$, $s_2^2 = 1.1429$, $s_3^2 = 1.6381$, $s_4^2 = 1.8381$, $s_5^2 = 1.4571$, and $s_6^2 = 1.6857$. *The variance associated with the total score is* $s_T^2 = 18.3810$. *We* calculated these variances using Excel and the appropriate variance function. Since the sum of the item-level variances is 8.1714, the alpha coefficient is calculated as follows:
- $\alpha = (6/(6-1)) \times (1 - 8.1714/18.3810) = 0.6665$.
- Is there a threshold that *α must exceed before we can conclude that the items are* internally consistent ? The answer is that there is no official and widely-accepted threshold.
- A rule of thumb that has been advocated in the literature (c.f. Nunnaly, 1978) is to require *α to equal 0.70 or exceed it before the items are considered internally* consistent. There are a few important comments about Cronbach's alpha that are worth mentioning: The *α coefficient as defined by equation 11.1 is expected to always fall between* 0 and 1. In reality, that will not always be the case, especially when the number subjects participating in the experiment is small. The alpha coefficient could indeed take a negative value. The cause of this odd situation is some negative between-item co-variances with a large absolute value. The only thing that is known with certainty is *α always being below 1.*

Validity

- The **validity** of a scale may be defined as the extent to which differences in observed scale scores reflect true differences among objects on the characteristic being measured, rather than systematic or random error. Perfect validity requires that there be no measurement error.

- **Content validity** is a subjective but systematic evaluation of how well the content of a scale represents the measurement task at hand.

- **Criterion validity** reflects whether a scale performs as expected in relation to other variables selected (criterion variables) as meaningful criteria.

Validity: Overview of Key Definitions

- **Validity (in general)**
  - The extent to which conclusions drawn from a study are true
- **Internal Validity**
  - When a researcher can clearly identify cause and effect relationships (i.e., there are no confounds)
- **External Validity**
  - The extent to which what you find in your study can be generalized to your target population
- **Construct Validity**
  - Extent to which your constructs of interest (e.g., sensation seeking) are accurately and completely identified (measured)
  - In other words, the extent to which you are actually measuring what you say you are measuring (your sensation seeking scale really does measure the true construct of sensation seeking)

Approaches to Assess Validity

- Content Validity (Face) - Representativeness of the sampling adequacy of the items contained in the measurement instrument.

- Criterion-Related Validity
    - Concurrent Validity: Two measures of the same construct are highly related. (Done in order to use a shortened version of the scale.
    - Predictive Validity: Ability of the scale to predict some outside criterion - generally some outcome, a current or future state of affairs
- Construct Validity
    - **Trait Validity**
        - **Reliability**
        - **Convergent Validity**: the extent to which one measure correlates highly with other methods designed to measure the same construct
        - **Discriminate Validity**: extent to which the measure is indeed novel and is simply <u>not</u> a reflection of some other variable. Quite simply, the measure should not correlate too highly with measures of different constructs.

    - **Nomological Validity:** Does the measure behave as expected. Are the relationships between the measure of interest with respect to sign and magnitude consistent with theory.

Validity

- **Construct validity** addresses the question of what construct or characteristic the scale is, in fact, measuring. Construct validity includes convergent, discriminant, and nomological validity.
- **Convergent validity** is the extent to which the scale correlates positively with other measures of the same construct.
- **Discriminant validity** is the extent to which a measure does not correlate with other constructs from which it is supposed to differ.
- **Nomological validity** is the extent to which the scale correlates in theoretically predicted ways with measures of different but related constructs.

Other Forms of Validity

- **Content Validity (Face Validity)**
  - Extent to which a measure is appropriate according to experts in the domain of interest
- **Concurrent Validity (Convergent Validity)**
  - Extent to which one measure of a construct overlaps with other similar measures of that construct
- **Discriminant Validity**
  - Extent to which a measure of one construct does not overlap with measures of different constructs
- **Predictive Validity**
  - Extent to which a measure of a construct can predict theoretically-relevant outcomes
- **Nomological Validity**
  - How a construct fits within a broader set of related constructs

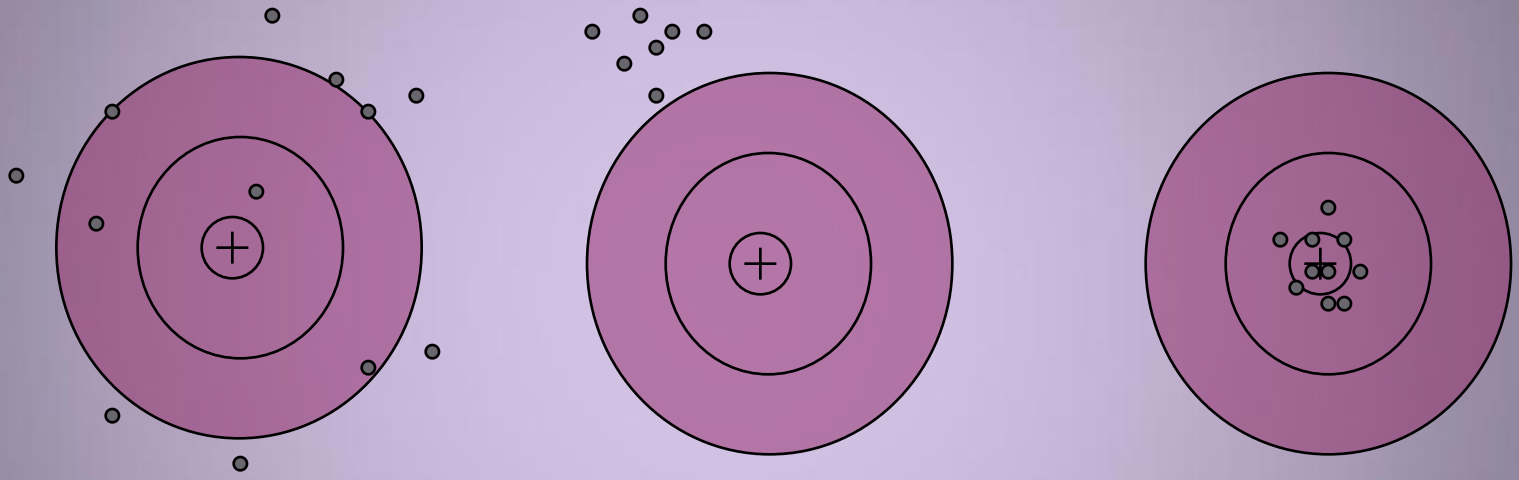
Relationship Between Reliability and Validity

- If a measure is perfectly valid, it is also perfectly reliable. In this case there is no random or systematic error.
- If a measure is unreliable, it cannot be perfectly valid, since at a minimum random error is present. Thus, unreliability implies invalidity.
- If a measure is perfectly reliable, it may or may not be perfectly valid, because systematic error may still be present.
- Reliability is a necessary, but not sufficient, condition for validity.

Validity and Reliability

- If a measure is valid, then it is reliable
- If it is not reliable, it can not be valid
- If it is reliable, it may or may not be valid
- Reliability can be more easily determined than validity

Reliability and Validity



Neither Reliable
Nor Valid

Reliable But
Not Valid

Reliable
And Valid

Example of low validity, high reliability

- Scale is perfectly accurate, but is capturing the wrong thing; for example, it measures consumers' interest in creative writing rather than preference for kinds of stationery.

Example of modest validity, low reliability

- Scale genuinely measures consumers' interest in kinds of stationery, but poorly worded items, sloppy administration, data entry errors lead to random errors in data
- Note that reliability sets an upper limit on validity -- a measure with a lot of errors is limited in how well it can capture a concept

Missing Data

- Try to guess from previous responses what value to insert (***not a good idea***)
- Substitute the average score for cases where data are present (***creates threats to validity***)
- Eliminate all cases for which any information is missing (***reduces the size of the usable data***)

THANK YOU
HAVE A NICE DAY