Definitions and Concepts

- **Independent variables** are variables or alternatives that are manipulated and whose effects are measured and compared, e.g., price levels.

- **Test units** are individuals, organizations, or other entities whose response to the independent variables or treatments is being examined, e.g., consumers or stores.

- **Dependent variables** are the variables which measure the effect of the independent variables on the test units, e.g., sales, profits, and market shares.

- **Extraneous variables** are all variables other than the independent variables that affect the response of the test units, e.g., store size, store location, and competitive effort (to be measured and controlled, even after randomization).
Validity in Experimentation

- **Internal validity** refers to whether the manipulation of the independent variables or treatments actually caused the observed effects on the dependent variables. Control of extraneous variables is a necessary condition for establishing internal validity.

- **External validity** refers to whether the cause-and-effect relationships found in the experiment can be generalized. To what populations, settings, times, independent variables, and dependent variables can the results be projected?
Measurement and Scaling: Fundamentals and Comparative Scaling

(Like in all sciences and fields, we need to measure concepts and variables, for both secondary data and primary data you collect!)
Chapter 8 Outline

1) Overview

2) Measurement and Scaling

3) Scales Characteristics
   i. Description
   ii. Order
   iii. Distance
   iv. Origin
Measurement and Scaling

**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.
Scale Characteristics

**Description**

By description, we mean the unique labels or descriptors that are used to designate each value of the scale. All scales possess description.

**Order**

By order, we mean the relative sizes or positions of the descriptors. Order is denoted by descriptors such as greater than, less than, and equal to.
Scale Characteristics

Distance
The characteristic of distance means that absolute differences between the scale descriptors are known and may be expressed in units (better be equal intervals).

Origin
The origin characteristic means that the scale has a unique or fixed beginning or true zero point.
Scaling involves creating a continuum upon which measured objects are located.

Consider an attitude scale from 1 to 100. Each respondent is assigned a number from 1 to 100, with 1 = Extremely Unfavorable, and 100 = Extremely Favorable. Measurement is the actual assignment of a number from 1 to 100 to each respondent. Scaling is the process of placing the respondents on a continuum with respect to their attitude toward department stores. (But people can seven bits of information at a time – thus 7 point scale is the most popular)
### Primary Scales of Measurement

<table>
<thead>
<tr>
<th>Scale</th>
<th>Fig. 8.1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nominal</strong></td>
<td>Numbers Assigned to Runners</td>
</tr>
<tr>
<td><strong>Ordinal</strong></td>
<td>Rank Order of Winners</td>
</tr>
<tr>
<td><strong>Interval</strong></td>
<td>Performance Rating on a 0 to 10 Scale</td>
</tr>
<tr>
<td><strong>Ratio</strong></td>
<td>Time to Finish in Seconds</td>
</tr>
</tbody>
</table>

#### Nominal
- Numbers assigned to runners: 7, 8, 3
- Finish lines indicating first, second, and third place.

#### Ordinal
- Rank order of winners: Third place, Second place, First place.
- Finish lines indicating first, second, and third place.

#### Interval
- Performance rating on a 0 to 10 scale: 8.2, 9.1, 9.6

#### Ratio
- Time to finish in seconds: 15.2, 14.1, 13.4
Primary Scales of Measurement
Nominal Scale

- The numbers serve only as labels or tags for identifying and classifying objects (no high vs. low, better vs. worse).

- 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.
### Illustration of Primary Scales of Measurement

#### Table 8.2

<table>
<thead>
<tr>
<th>No.</th>
<th>Store</th>
<th>Nominal Scale</th>
<th>Ordinal Scale</th>
<th>Interval Scale</th>
<th>Ratio Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Preference</td>
<td>Rankings</td>
<td>Preference Ratings</td>
<td>$ spent last 3 months</td>
</tr>
<tr>
<td>1.</td>
<td>Parisian</td>
<td>7</td>
<td>79</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>2.</td>
<td>Macy’s</td>
<td>2</td>
<td>25</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>3.</td>
<td>Kmart</td>
<td>8</td>
<td>82</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>4.</td>
<td>Kohl’s</td>
<td>3</td>
<td>30</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>5.</td>
<td>J.C. Penney</td>
<td>1</td>
<td>10</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>6.</td>
<td>Neiman Marcus</td>
<td>5</td>
<td>53</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>7.</td>
<td>Marshalls</td>
<td>9</td>
<td>95</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>8.</td>
<td>Saks Fifth Avenue</td>
<td>6</td>
<td>61</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>9.</td>
<td>Sears</td>
<td>4</td>
<td>45</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>10.</td>
<td>Wal-Mart</td>
<td>10</td>
<td>115</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>
Primary Scales of Measurement

Ordinal Scale

- 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 (1st, 2nd....).
- 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 (such as top ten percentile or the 1st decile).
Primary Scales of Measurement

Interval Scale

• 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 (can run correlations and regression!).

• 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, and in addition the arithmetic mean, standard deviation, and other statistics commonly used in marketing research.
Primary Scales of Measurement
Ratio Scale

- 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.
# Primary Scales of Measurement

## Table 8.1

<table>
<thead>
<tr>
<th>Scale</th>
<th>Basic Characteristics</th>
<th>Common Examples</th>
<th>Marketing Examples</th>
<th>Permissible Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Descriptive</strong></td>
</tr>
<tr>
<td><strong>Nominal</strong></td>
<td>Numbers identify &amp; classify objects</td>
<td>Social Security nos., numbering of football players</td>
<td>Brand nos., store types</td>
<td>Percentages, mode</td>
</tr>
<tr>
<td><strong>Ordinal</strong></td>
<td>Nos. indicate the relative positions of objects but not the magnitude of differences between them</td>
<td>Quality rankings, rankings of teams in a tournament</td>
<td>Preference rankings, market position, social class</td>
<td>Percentile, median</td>
</tr>
<tr>
<td><strong>Interval</strong></td>
<td>Differences between objects</td>
<td>Temperature (Fahrenheit)</td>
<td>Attitudes, opinions, index</td>
<td>Range, mean, standard</td>
</tr>
<tr>
<td><strong>Ratio</strong></td>
<td>Zero point is fixed, ratios of scale values can be compared</td>
<td>Length, weight</td>
<td>Age, sales, income, costs</td>
<td>Geometric mean, harmonic mean</td>
</tr>
</tbody>
</table>
A Classification of Scaling Techniques

Figure 8.2

- Scaling Techniques
  - Comparative Scales
    - Paired Comparison
    - Rank Order
    - Constant Sum
  - Noncomparative Scales
    - Q-Sort and Other Procedures
    - Continuous Rating Scales
    - Itemized Rating Scales
      - Likert
      - Semantic Differential
      - Stapel
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 **noncomparative 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 (relative to or in comparison with....).

• 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.
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.
### 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:

<table>
<thead>
<tr>
<th></th>
<th>Jhirmack</th>
<th>Finesse</th>
<th>Vidal Sassoon</th>
<th>Head &amp; Shoulders</th>
<th>Pert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jhirmack</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Finesse</td>
<td>1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Vidal Sassoon</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Head &amp; Shoulders</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pert</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Number of Times Preferred&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>a</sup>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. <sup>b</sup>The number of times a brand was preferred is obtained by summing the 1s in each column.
Paired Comparison Selling

The most common method of taste testing is paired comparison. The consumer is asked to sample two different products and select the one with the most appealing taste. The test is done in private and a minimum of 1,000 responses is considered an adequate sample. A blind taste test for a soft drink, where imagery, self-perception and brand reputation are very important factors in the consumer’s purchasing decision, may not be a good indicator of performance in the marketplace. The introduction of New Coke illustrates this point. New Coke was heavily favored in blind paired comparison taste tests, but its introduction was less than successful, because image plays a major role in the purchase of Coke.

A paired comparison taste test
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

Fig. 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

Fig. 8.4 cont.

Form

<table>
<thead>
<tr>
<th>Brand</th>
<th>Rank Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Crest</td>
<td></td>
</tr>
<tr>
<td>2. Colgate</td>
<td></td>
</tr>
<tr>
<td>3. Aim</td>
<td></td>
</tr>
<tr>
<td>4. Gleem</td>
<td></td>
</tr>
<tr>
<td>5. Sensodyne</td>
<td></td>
</tr>
<tr>
<td>6. Ultra Brite</td>
<td></td>
</tr>
<tr>
<td>7. Close Up</td>
<td></td>
</tr>
<tr>
<td>8. Pepsodent</td>
<td></td>
</tr>
<tr>
<td>9. Plus White</td>
<td></td>
</tr>
<tr>
<td>10. Stripe</td>
<td></td>
</tr>
</tbody>
</table>
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

Fig. 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

#### Fig. 8.5 cont.

#### Form

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Segment I</th>
<th>Segment II</th>
<th>Segment III</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mildness</td>
<td>8</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2. Lather</td>
<td>2</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>3. Shrinkage</td>
<td>3</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>4. Price</td>
<td>53</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>5. Fragrance</td>
<td>9</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>6. Packaging</td>
<td>7</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>7. Moisturizing</td>
<td>5</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>8. Cleaning Power</td>
<td>13</td>
<td>60</td>
<td>15</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>
Chapter 9 Outline

5) Noncomparative Itemized Rating Scale Decisions
   i. Number of Scale Categories
   ii. Balanced Vs. Unbalanced Scales
   iii. Odd or Even Number of Categories
   iv. Forced Vs. Non-forced Scales
   v. Nature and Degree of Verbal Description
   vi. Physical Form or Configuration

6) Multi-item Scales
Noncomparative Scaling Techniques

• Respondents evaluate only one object at a time, and for this reason non-comparative scales are often referred to as monadic scales.

• Non-comparative techniques consist of continuous and itemized rating scales.
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
A relatively new research tool, the perception analyzer, provides continuous measurement of “gut reaction.” A group of up to 400 respondents is presented with TV or radio spots or advertising copy. The measuring device consists of a dial that contains a 100-point range. Each participant is given a dial and instructed to continuously record his or her reaction to the material being tested.

As the respondents turn the dials, the information is fed to a computer, which tabulates second-by-second response profiles. As the results are recorded by the computer, they are superimposed on a video screen, enabling the researcher to view the respondents' scores immediately. The responses are also stored in a permanent data file for use in further analysis. The response scores can be broken down by categories, such as age, income, sex, or product usage.
Itemized Rating Scales

• The respondents are provided with a scale that has a number or brief description associated with each category.

• The categories are ordered in terms of scale position, and the respondents are required to select the specified category that best describes the object being rated.

• The commonly used itemized rating scales are the Likert, semantic differential, and Stapel scales.
The **Likert scale** requires the respondents to indicate a degree of agreement or disagreement with each of a series of statements about the stimulus objects.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sears sells high-quality merchandise.</td>
<td>1</td>
<td>2X</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2. Sears has poor in-store service.</td>
<td>1</td>
<td>2X</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3. I like to shop at Sears.</td>
<td>1</td>
<td>2</td>
<td>3X</td>
<td>4</td>
</tr>
</tbody>
</table>

- The analysis can be conducted on an item-by-item basis (profile analysis), or a total (summated) score can be calculated.

- When arriving at a total score, the categories assigned to the negative statements by the respondents should be scored by reversing the scale.
The **semantic differential** is a seven-point rating scale with end points associated with **bipolar labels** that have semantic meaning.

**SEARS IS:**
- Powerful --:--:--:--:--:--:--:--: Weak
- Unreliable --:--:--:--:--:--:--:X--: Reliable
- Modern --:--:--:--:--:--:--:--:X: Old-fashioned

- The negative adjective or phrase sometimes appears at the left side of the scale and sometimes at the right.
- This controls the tendency of some respondents, particularly those with very positive or very negative attitudes, to mark the right- or left-hand sides without reading the labels.
- Individual items on a semantic differential scale may be scored on either a -3 to +3 or a 1 to 7 scale.

1) Rugged :---:---:---:---:---:---:---: Delicate
2) Excitable :---:---:---:---:---:---:---: Calm
3) Uncomfortable :---:---:---:---:---:---:---: Comfortable
4) Dominating :---:---:---:---:---:---:---: Submissive
5) Thrifty :---:---:---:---:---:---:---: Indulgent
6) Pleasant :---:---:---:---:---:---:---: Unpleasant
7) Contemporary :---:---:---:---:---:---:---: Obsolete
8) Organized :---:---:---:---:---:---:---: Unorganized
9) Rational :---:---:---:---:---:---:---: Emotional
10) Youthful :---:---:---:---:---:---:---: Mature
11) Formal :---:---:---:---:---:---:---: Informal
12) Orthodox :---:---:---:---:---:---:---: Liberal
13) Complex :---:---:---:---:---:---:---: Simple
14) Colorless :---:---:---:---:---:---:---: Colorful
15) Modest :---:---:---:---:---:---:---: Vain
The **Stapel scale** is a unipolar rating scale with ten categories numbered from -5 to +5, without a neutral point (zero). This scale is usually presented vertically.

<table>
<thead>
<tr>
<th>SEARS</th>
<th>SEARS</th>
</tr>
</thead>
<tbody>
<tr>
<td>+5</td>
<td>+5</td>
</tr>
<tr>
<td>+4</td>
<td>+4</td>
</tr>
<tr>
<td>+3</td>
<td>+3</td>
</tr>
<tr>
<td>+2</td>
<td>+2X</td>
</tr>
<tr>
<td>+1</td>
<td>+1</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>-3</td>
<td>-3</td>
</tr>
<tr>
<td>-4X</td>
<td>-4</td>
</tr>
<tr>
<td>-5</td>
<td>-5</td>
</tr>
</tbody>
</table>

The data obtained by using a Stapel scale can be analyzed in the same way as semantic differential data.
# Basic Noncomparative Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Basic Characteristics</th>
<th>Examples</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous Rating Scale</td>
<td>Place a mark on a continuous line</td>
<td>Reaction to TV commercials</td>
<td>Easy to construct</td>
<td>Scoring can be cumbersome unless computerized</td>
</tr>
<tr>
<td><strong>Itemized Rating Scales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likert Scale</td>
<td>Degrees of agreement on a 1 (strongly disagree) to 5 (strongly agree) scale</td>
<td>Measurement of attitudes</td>
<td>Easy to construct, administer, and understand</td>
<td>More time-consuming</td>
</tr>
<tr>
<td>Semantic Differential</td>
<td>Seven-point scale with bipolar labels</td>
<td>Brand, product, and company images</td>
<td>Versatile</td>
<td>Controversy as to whether the data are interval</td>
</tr>
<tr>
<td>Stapel Scale</td>
<td>Unipolar ten-point scale, - 5 to +5, without a neutral point (zero)</td>
<td>Measurement of attitudes and images</td>
<td>Easy to construct, administer over telephone</td>
<td>Confusing and difficult to apply</td>
</tr>
</tbody>
</table>
## Summary of Itemized Scale Decisions

### Table 9.2

<table>
<thead>
<tr>
<th>1) <strong>Number of categories</strong></th>
<th>Although there is no single, optimal number, traditional guidelines suggest that there should be between five and nine categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>2) <strong>Balanced vs. unbalanced</strong></td>
<td>In general, the scale should be balanced to obtain objective data</td>
</tr>
<tr>
<td>3) <strong>Odd/even no. of categories</strong></td>
<td>If a neutral or indifferent scale response is possible for at least some respondents, an odd number of categories should be used</td>
</tr>
<tr>
<td>4) <strong>Forced vs. non-forced</strong></td>
<td>In situations where the respondents are expected to have no opinion, the accuracy of the data may be improved by a non-forced scale</td>
</tr>
<tr>
<td>5) <strong>Verbal description</strong></td>
<td>An argument can be made for labeling all or many scale categories. The category descriptions should be located as close to the response categories as possible</td>
</tr>
<tr>
<td>6) <strong>Physical form</strong></td>
<td>A number of options should be tried and the best selected</td>
</tr>
</tbody>
</table>
Balanced and Unbalanced Scales

Fig. 9.1

<table>
<thead>
<tr>
<th>Balanced Scale</th>
<th>Unbalanced Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jovan Musk for Men is:</td>
<td>Jovan Musk for Men is:</td>
</tr>
<tr>
<td>Extremely good</td>
<td>Extremely good</td>
</tr>
<tr>
<td>Very good</td>
<td>Very good</td>
</tr>
<tr>
<td>Good</td>
<td>Good</td>
</tr>
<tr>
<td>Bad</td>
<td>Somewhat good</td>
</tr>
<tr>
<td>Very bad</td>
<td>Bad</td>
</tr>
<tr>
<td>Extremely bad</td>
<td>Very bad</td>
</tr>
</tbody>
</table>

Copyright © 2010 Pearson Education, Inc. publishing as Prentice Hall
Rating Scale Configurations

Fig. 9.2

Cheer detergent is:

1) Very harsh --- --- --- --- --- --- --- Very gentle
2) Very harsh 1 2 3 4 5 6 7 Very gentle
3) Very harsh
   .
   . Neither harsh nor gentle
   .
   . Very gentle
4) Very harsh Harsh Somewhat harsh Neither harsh nor gentle Somewhat gentle Gentle Very gentle
5) -3 -2 -1 0 +1 +2 +3
   Very harsh Neither harsh nor gentle Very gentle
Some Unique Rating Scale Configurations

**Thermometer Scale**
Instructions: Please indicate how much you like McDonald’s hamburgers by coloring in the thermometer. Start at the bottom and color up to the temperature level that best indicates how strong your preference is.

**Smiling Face Scale**
Instructions: Please point to the face that shows how much you like the Barbie Doll. If you do not like the Barbie Doll at all, you would point to Face 1. If you liked it very much, you would point to Face 5.
Development of a Multi-item Scale

Fig. 9.4

1. **Develop Theory**
2. **Generate Initial Pool of Items: Theory, Secondary Data, and Qualitative Research**
3. **Select a Reduced Set of Items Based on Qualitative Judgment**
4. **Collect Data from a Large Pretest Sample**
5. **Statistical Analysis**
6. **Develop Purified Scale**
7. **Collect More Data from a Different Sample**
8. **Evaluate Scale Reliability, Validity, and Generalizability**
9. **Final Scale**
Scale Evaluation

Fig. 9.5

Scale Evaluation

- Reliability
  - Test/Retest
  - Alternative Forms
  - Internal Consistency

- Validity
  - Content
  - Criterion
  - Construct
    - Convergent
    - Discriminant
    - Nomological

- Generalizability
Measurement Accuracy

The **true score model** provides a framework for understanding the accuracy of measurement.

\[ X_O = X_T + X_S + X_R \]

where

\[ X_O = \text{the observed score or measurement} \]
\[ X_T = \text{the true score of the characteristic} \]
\[ X_S = \text{systematic error} \]
\[ X_R = \text{random error} \]
Potential Sources of Error on Measurement

1) Other relatively stable characteristics of the individual that influence the test score, such as intelligence, social desirability, and education.

2) Short-term or transient personal factors, such as health, emotions, and fatigue.

3) Situational factors, such as the presence of other people, noise, and distractions.

4) Sampling of items included in the scale: addition, deletion, or changes in the scale items.

5) Lack of clarity of the scale, including the instructions or the items themselves.

6) Mechanical factors, such as poor printing, overcrowding items in the questionnaire, and poor design.

7) Administration of the scale, such as differences among interviewers.

8) Analysis factors, such as differences in scoring and statistical analysis.
Reliability

- **Reliability** can be defined as the extent to which measures are free from random error, $X_R$. If $X_R = 0$, the measure is perfectly reliable.

- 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.

- 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.
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 \( (X_O = X_T, X_R = 0, X_S = 0) \).

• **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

- **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.
Relationship Between Reliability and Validity

• If a measure is perfectly valid, it is also perfectly reliable. In this case $X_O = X_T$, $X_R = 0$, and $X_S = 0$.

• If a measure is unreliable, it cannot be perfectly valid, since at a minimum $X_O = X_T + X_R$. Furthermore, systematic error may also be present, i.e., $X_S \neq 0$. 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 ($X_O = X_T + X_S$).

• Reliability is a necessary, but not sufficient, condition for validity.
1. Based on your understanding of the research problems, please list out all the major concepts/variables you need to know and define each one of them. A single indicator or multiple items (unobserved, abstract or composite concepts)?

Remember to distinguish symptoms of problems from the underlying causes or issues.

Remember to translate the management decision problems into research problems!
2. Try to draw a flow-chart model to define the relationships among them: from left/up (independent vars) to right/down (dependent vars), direct causal relationship (correlational, causal, positive or negative, linear vs. nonlinear, mediating, moderating)

Where do theories or framework come from?

From topic and knowledge areas in your fields

From literature review.
3. For each of the variable, please indicate the source of data (primary or secondary based on the research design in GP3), the scale used for all the variables, and their items.

Remember definition (validity) and reliability issues.
Group Project 4: Concepts, (Theory) Framework, and scales

1. Based on your understanding of the project and research problems, please list out all the major concepts/variables you need to know and define each one of them. A single indicator or multiple items (unobserved, abstract or composite concepts)?
2. Try to draw a flow-chart model to define the relationships among them: from left/up (independent vars) to right/down (dependent vars), direct causal relationship (positive or negative, mediating, moderating)
3. For each of the variable, please indicate the source of data (primary or secondary based on the research design in GP3), the scale used for all the variables, and their items.