Chapter 10

Measurement and Scaling: Noncomparative Scaling Techniques
Figure 10.1 Relationship of Noncomparative Scaling to the Previous Chapters and the Marketing Research Process

Focus of this Chapter
- Continuous Rating Scales
- Itemized Rating Scales

Relationship to Previous Chapters
- Research Design Components (Chapter 3)
- Basic Types of Scales (Chapter 9)

Relationship to Marketing Research Process
1. Problem Definition
2. Approach to Problem
3. Research Design
4. Field Work
5. Data Preparation and Analysis
6. Report Preparation and Presentation

Copyright © 2011 Pearson Education, Inc.  Chapter 10 - 2
Figure 10.2  Noncomparative Scaling Techniques: An Overview
Opening Vignette

Noncomparative Scaling Techniques (Fig 10.3) (Table 10.1)

Continuous Rating Scale

Itemized Rating Scales

- Likert
- Stapel
- Semantic Differential

Rating Scale Decisions and Multi-item Scales (Figs 10.4, 10.5 & 10.6) (Tables 10.2 & 10.3)

Scale Evaluation (Fig 10.7)

Choosing a Scaling Technique

Application to Contemporary Issues (Fig 10.8)

- International
- Social Media
- Ethics

Copyright © 2011 Pearson Education, Inc. Chapter 10 - 4
Noncomparative Scaling Techniques

- Respondents evaluate only one object at a time, and for this reason noncomparative scales are often referred to as monadic scales.
- Noncomparative techniques consist of continuous and itemized rating scales.
Figure 10.3
A Classification of Noncomparative Rating Scales

Noncomparative Rating Scales

Continuous Rating Scales

Semantic Differential

Itemized Rating Scales

Stapel

Likert
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
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.
## Likert Scale

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>Statement</th>
<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>
<td>5</td>
</tr>
<tr>
<td>2. Sears has poor in-store service.</td>
<td>1</td>
<td>2X</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3. I like to shop at Sears.</td>
<td>1</td>
<td>2</td>
<td>3X</td>
<td>4</td>
<td>5</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.

<table>
<thead>
<tr>
<th>SEARS is:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powerful --:--:--:--:X--:--: Weak</td>
</tr>
<tr>
<td>Unreliable --:--:--:--:--X--: Reliable</td>
</tr>
<tr>
<td>Modern --:--:--:--:--:--:X: Old-fashioned</td>
</tr>
</tbody>
</table>

- 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.
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></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>High Quality</td>
<td></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.
<table>
<thead>
<tr>
<th>Scale</th>
<th>Basic Characteristics</th>
<th>Examples</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Continuous Rating Scale</strong></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><strong>Likert Scale</strong></td>
<td>Degree 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><strong>Semantic Differential</strong></td>
<td>Seven-point scale with bipolar labels</td>
<td>Brand, product, and company images</td>
<td>Versatile</td>
<td>Difficult to construct bipolar adjectives</td>
</tr>
<tr>
<td><strong>Stapel Scale</strong></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 and administer over telephone</td>
<td>Confusing and difficult to apply</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. <strong>Number of categories</strong></td>
<td>While there is no single, optimal number, traditional guidelines suggest that there should be between five and nine categories.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. <strong>Balanced vs. unbalanced</strong></td>
<td>In general, the scale should be balanced to obtain objective data.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. <strong>Odd or even number of Categories</strong></td>
<td>If a neutral or indifferent scale response is possible for at least some of the respondents, an odd number of categories should be used.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Table 10.2**  
Summary of Itemized Rating Scale Decisions (Cont.) |   |
<p>| <strong>4. Forced versus nonforced</strong> | In situations where the respondents are expected to have no opinion, the accuracy of data may be improved by a nonforced scale. |
| <strong>5. Verbal description</strong> | 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. |
| <strong>6. Physical form</strong> | A number of options should be tried and the best one selected. |</p>
<table>
<thead>
<tr>
<th>Balanced Scale</th>
<th>Unbalanced Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surfing the Internet is</strong></td>
<td><strong>Surfing the Internet is</strong></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>

Figure 10.4 Balanced and Unbalanced Scales
A variety of scale configurations may be employed to measure the comfort of Nike shoes. Some examples include:

Nike shoes are:
1) **Place an “X” on one of the blank spaces…**
   - Very
   - Uncomfortable

2) **Circle the number…**
   - Very
   - Uncomfortable
   - 1 2 3 4 5 6 7

3) **Place an “X” on one of the blank spaces…**
   - Very Uncomfortable
   - Uncomfortable
   - Neither Uncomfortable nor Comfortable
   - Comfortable
   - Very Comfortable
### Figure 10.5  Rating Scale Configurations (Cont.)

<table>
<thead>
<tr>
<th>4)</th>
<th>Very Uncomfortable</th>
<th>Uncomfortable</th>
<th>Somewhat Uncomfortable</th>
<th>Neither Comfortable nor Uncomfortable</th>
<th>Somewhat Comfortable</th>
<th>Comfortable</th>
<th>Very Comfortable</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>5)</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very Uncomfortable</td>
<td>Neither Comfortable nor Uncomfortable</td>
<td>Very Comfortable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

Form:

![Thermometer Scale](image)

<table>
<thead>
<tr>
<th>Temperature Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Like very much</td>
</tr>
<tr>
<td>75</td>
<td>Like much</td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Dislike very much</td>
</tr>
</tbody>
</table>

**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.

Form:

![Smiling Face Scale](image)

<table>
<thead>
<tr>
<th>Face</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Dislike</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Like very much</td>
</tr>
<tr>
<td>5</td>
<td>Like much</td>
</tr>
<tr>
<td>Construct</td>
<td>Scale Descriptors</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>Attitude</td>
<td>Very Bad, Bad, Neither Bad nor Good, Good, Good, Very Good</td>
</tr>
<tr>
<td>Importance</td>
<td>Not at All Important, Not Important, Neutral, Important, Very Important</td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Very Dissatisfied, Dissatisfied, Neither Dissatisfied nor Satisfied, Satisfied, Very Satisfied</td>
</tr>
<tr>
<td>Purchase Frequency</td>
<td>Never, Rarely, Sometimes, Often, Very Often</td>
</tr>
</tbody>
</table>
Figure 10.6 Developing a Multi-Item Scale

- Develop the Constraint
- Develop a Theoretical Definition
- Develop an Operational Definition
- **Develop a Multi-Item Scale**
  - Generate a Pool of Scale Items
  - Reduce the Pool of Items Based on Judgment
  - Collect Data
  - Purify the Scale Based on Statistical Analysis
- Evaluate Scale Reliability and Validity
- **Apply the Scale and Accumulate Research Findings**

Copyright © 2011 Pearson Education, Inc.  Chapter 10 - 20
Figure 10.7 Scale Evaluation

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

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

Scale Evaluation
Total Measurement Error

**Total measurement error** is the sum of systematic error and random error.

Total measurement error = Systematic error + Random error

- **Systematic error** affects the measurement in a constant way, that is, in the same way each time the measurement is made.
- **Random error**, in contrast, arises from random changes and has a different effect each time the measurement is made.
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 (Cont.)

- **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.

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

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

- **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.
Validity (Cont.)

- **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, 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.
Chapter 10 - Noncomparative Scales

Figure 10.8 A Concept Map for Noncomparative Scales

Basic Noncomparative Scales

- Continuous Rating Scales
  - basic characteristic
  - Place a Mark on a Continuous Line
    - example: Reaction to TV Commercials
- Itemized Rating Scales
  - may be
    - Likert
      - basic characteristic: Degree of Agreement on a 1-5 Scales
        - example: Measurement of Attitudes
    - Semantic Differential
      - basic characteristic: 7-point scale with Bipolar Labels
        - example: Brand Image Measurement
    - Stapel
      - basic characteristic: 10-point scale, -5 to +5, without a neutral point (Zero)
        - example: Measurement of Attitudes and Images

Copyright © 2011 Pearson Education, Inc. Chapter 10 - 29
Pan-cultural scales, designed to be free of cultural biases, are used in international research.

An approach to developing pan-cultural scales is to use descriptors the respondents create themselves. The end descriptors used to anchor the scale are particularly prone to different interpretations.

Additionally, the scale numbering may have different meanings. In such cases, it might be desirable to avoid numbers and to just use boxes that respondents can check.
International Marketing Research (Cont.)

- Of the scaling techniques considered, the semantic differential has been applied with the greatest consistency in results across countries.
- Rating scales also are used to construct indexes used to make global comparisons.
Marketing Research & Social Media

- All the noncomparative scales can be easily implemented in social media.
- An analysis of social media content can provide guidance on whether continuous or itemized rating scales should be used. It can also aid the researcher in making appropriate rating scale decisions.
- It is feasible to employ more than one scaling method to measure a given construct.
Ethics in Marketing Research

- The researcher’s should not deliberately bias the results by building that bias into noncomparative scales.
- The researcher should establish the reliability and validity of scales.
- The researcher has a responsibility to both the client and respondents to ensure the applicability and usefulness of the scales.
Acronym: Rating

The rating scale decisions may be described by the acronym RATING:

R esponse option: forced versus nonforced
A ttractive versus unattractive number of categories: balanced versus unbalanced
T otal number of categories
I mpartial or neutral category: odd versus even number of categories
N ature and degree of verbal description
G raphics: physical form and configuration