



## Chapter Nine

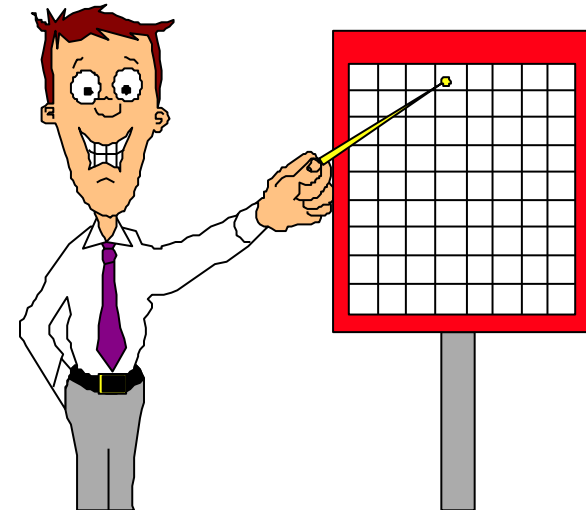
---

# Measurement and Scaling: Noncomparative Scaling Techniques

# Chapter Outline

---

- 1) Overview
- 2) Noncomparative Scaling Techniques
- 3) Continuous Rating Scale
- 4) Itemized Rating Scale
  - i. Likert Scale
  - ii. Semantic Differential Scale
  - iii. Stapel Scale



# Chapter 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



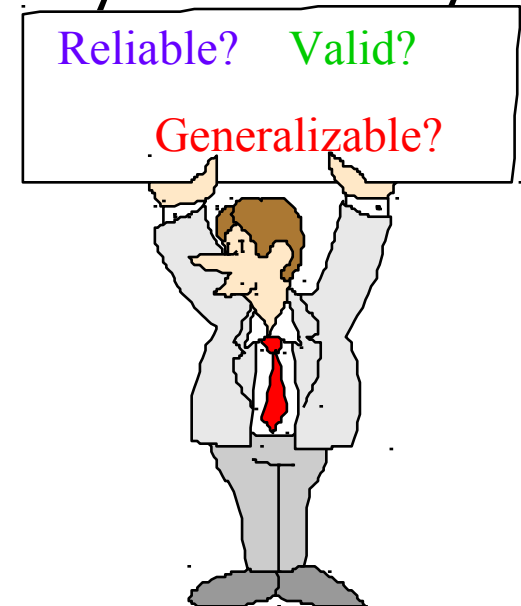
# Chapter Outline

## 7) Scale Evaluation

- i. Measurement Accuracy
- ii. Reliability
- iii. Validity
- iv. Relationship between Reliability and Validity
- v. Generalizability

## 8) Choosing a Scaling Technique

## 9) Mathematically Derived Scales



# Chapter Outline

---

- 10) International Marketing Research
- 11) Ethics in Marketing Research
- 12) Internet and Computer Applications
- 13) Focus on Burke
- 14) Summary
- 15) Key Terms and Concepts





# Noncomparative Scaling Techniques

---

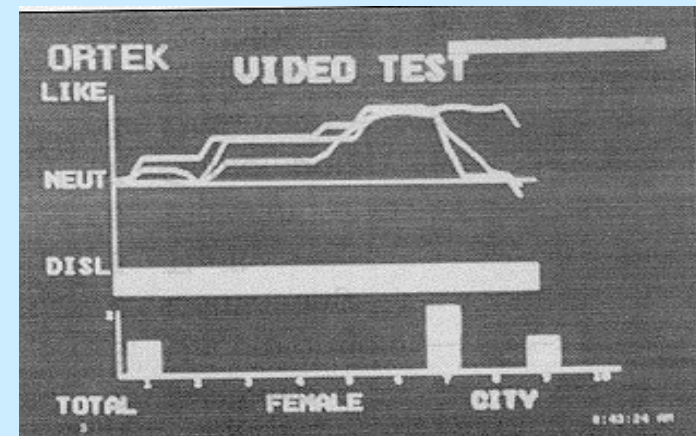
- Respondents evaluate only one object at a time, and for this reason noncomparative scales are often referred to as monadic scales.
- Noncomparative techniques consist of continuous and itemized rating scales.



# RATE: Rapid Analysis and Testing Environment

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.



# 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.

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	
1. Sears sells high quality merchandise.	1	2X	3	4	5	
2. Sears has poor in-store service.		1	2X	3	4	5
3. I like to shop at Sears.	1	2	3X	4	5	

- 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.



# Semantic Differential Scale

---

The **semantic differential** is a seven-point rating scale with end points associated with bipolar labels that have semantic meaning.

SEARS IS:

Powerful    --:--:--:--:-X-:--:--: 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.

# A Semantic Differential Scale for Measuring Self-Concepts, Person Concepts, and Product Concepts

- 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





# Stapel 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.

SEARS	
+5	+5
+4	+4
+3	+3
+2	+2X
+1	+1
HIGH QUALITY	POOR SERVICE
-1	-1
-2	-2
-3	-3
-4X	-4
-5	-5

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



# Basic Noncomparative Scales

Table 9.1

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

# Summary of Itemized Scale Decisions

Table 9.2

## 1) Number of categories

**Although there is no single, optimal number, traditional guidelines suggest that there should be between five and nine categories**

## 2) Balanced vs. unbalanced

**In general, the scale should be balanced to obtain objective data**

## 3) Odd/even no. of categories

**If a neutral or indifferent scale response is possible from at least some of the respondents, an odd number of categories should be used**

## 4) Forced vs. non-forced

**In situations where the respondents are expected to have no opinion, the accuracy of the data may be improved by a non-forced scale**

## 5) Verbal description

**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**

## 6) Physical form

**A number of options should be tried and the best selected**



# Balanced and Unbalanced Scales

Figure 9.1

## Balanced Scale

Jovan Musk for Men is

Extremely good \_\_\_\_\_

Very good \_\_\_\_\_

Good \_\_\_\_\_

Bad \_\_\_\_\_

Very bad \_\_\_\_\_

Extremely bad \_\_\_\_\_

## Unbalanced Scale

Jovan Musk for Men is

Extremely good \_\_\_\_\_

Very good \_\_\_\_\_

Good \_\_\_\_\_

Somewhat good \_\_\_\_\_

Bad \_\_\_\_\_

Very bad \_\_\_\_\_





# Rating Scale Configurations

Figure 9.2

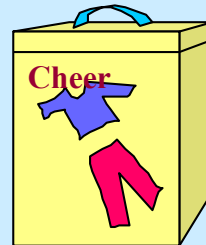
A variety of scale configurations may be employed to measure the gentleness of Cheer detergent. Some examples include:

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
<input type="text" value="-3"/>	<input type="text" value="-2"/>	<input type="text" value="-1"/>	<input type="text" value="0"/>	<input type="text" value="+1"/>	<input type="text" value="+2"/>	<input type="text" value="+3"/>

5) 

Very harsh			Neither harsh nor gentle			Very gentle
------------	--	--	--------------------------	--	--	-------------

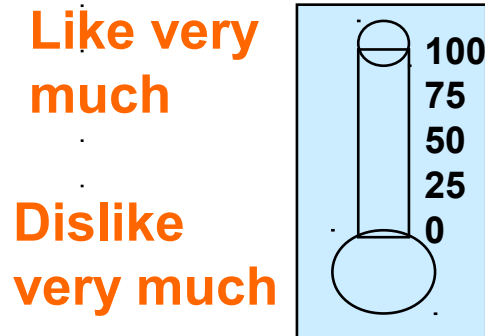
# Some Unique Rating Scale Configurations

Figure 9.3

## 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.

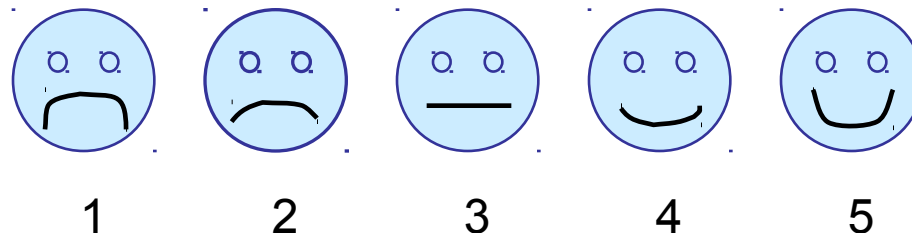
Form:



## Smiling Face Scale

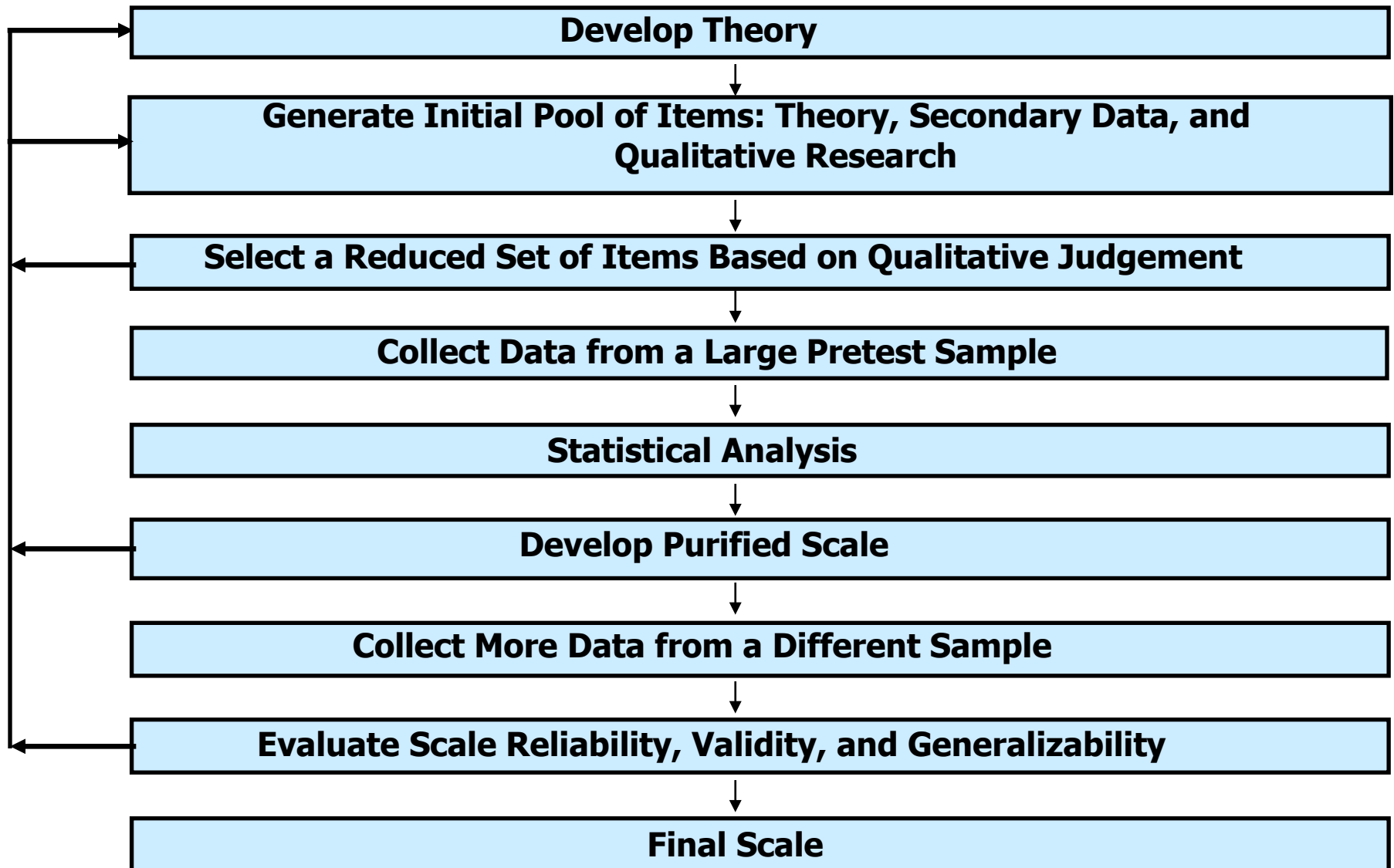
Instructions: Please point to the face that shows how much you like the Barbie Doll. If you do not like the Barbie Doll at all, you would point to Face 1. If you liked it very much, you would point to Face 5.

Form:



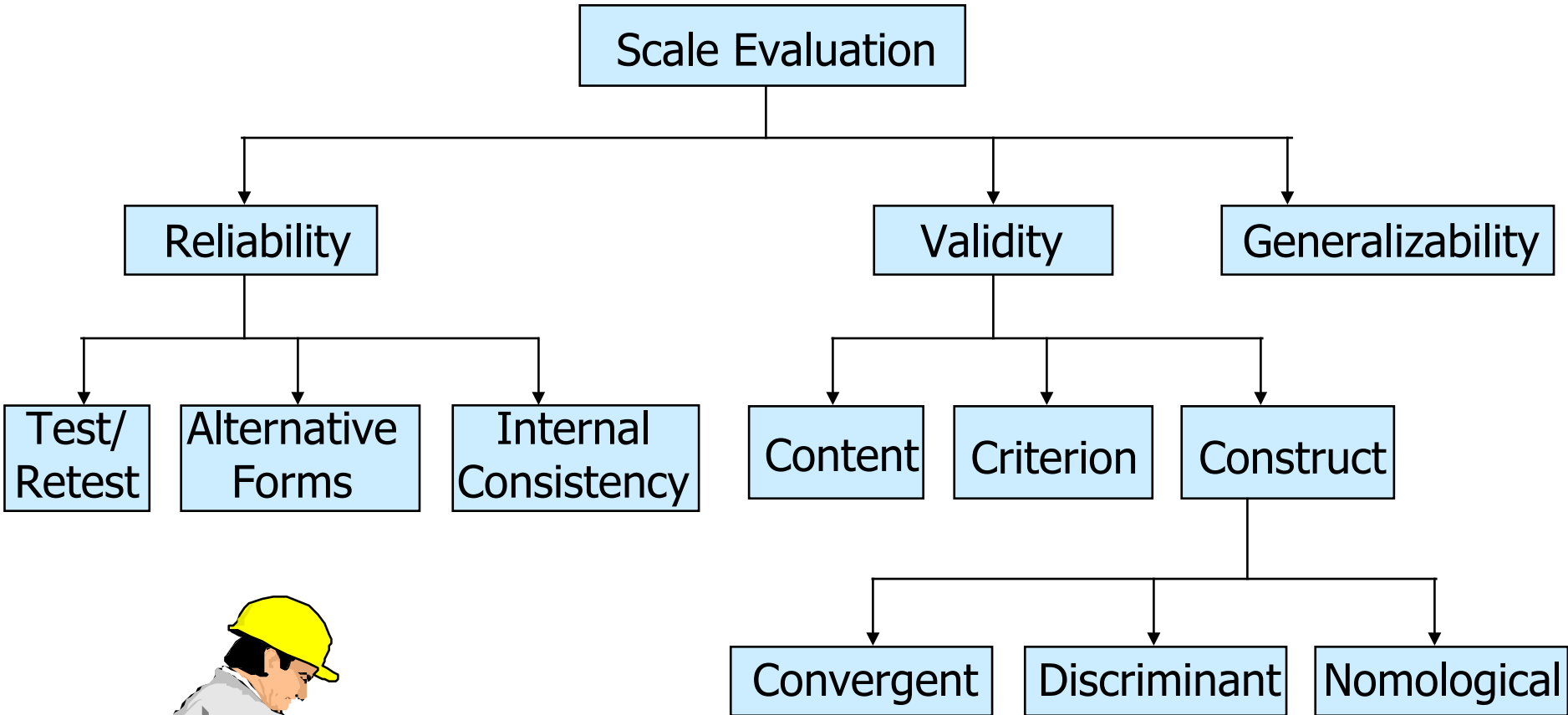
# Development of a Multi-item Scale

Figure 9.4



# Scale Evaluation

Figure 9.5





# Measurement Accuracy

---

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

$$X_O = X_T + X_S + X_R$$

where

$X_O$  = the observed score or measurement

$X_T$  = the true score of the characteristic

$X_S$  = systematic error

$X_R$  = random error



# Potential Sources of Error on Measurement

---

Figure 9.6

- 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.