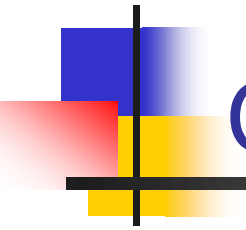


# Chapter Eight

## Measurement and Scaling: Fundamentals and Comparative Scaling

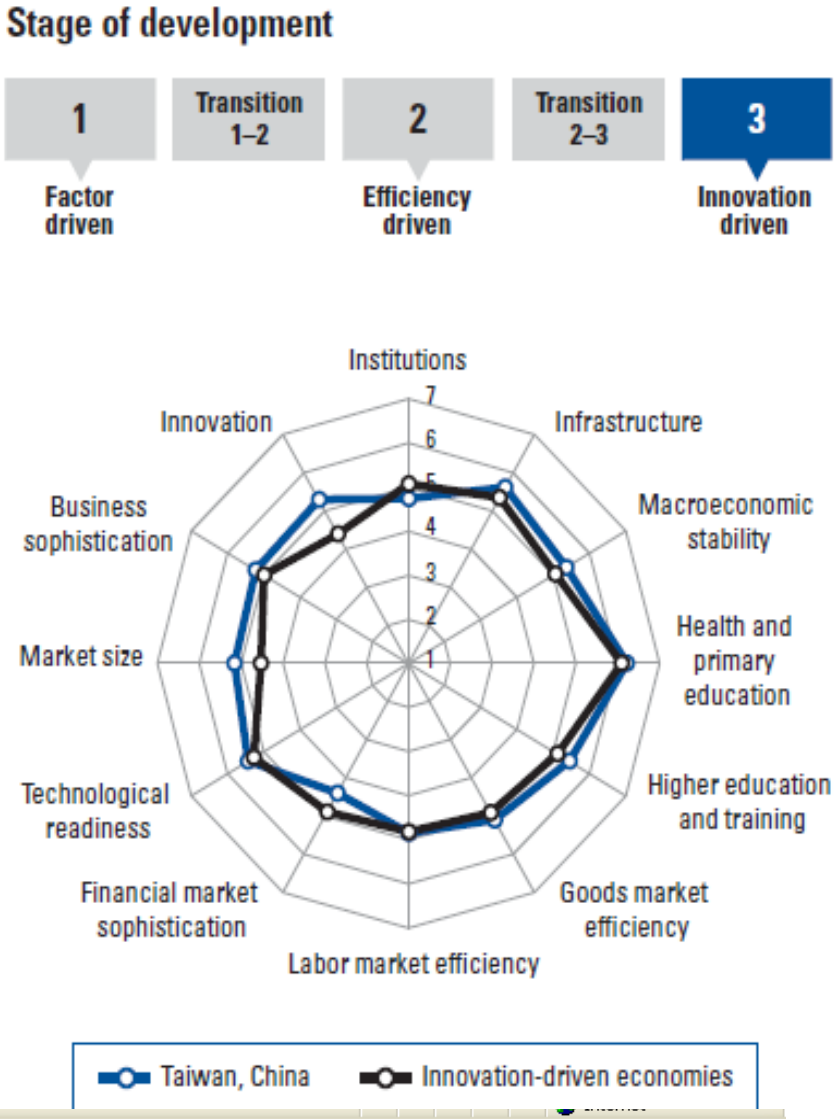
# If you can not measure, you can not manage.
## - Total Quality Management

- Examples: mate selection, buying house, buying stocks, faculty member's promotion, quality improvement, consumer satisfaction, suicide tendency, criminal DNA, IQ, EQ, talented employee, nostalgia, CSR, virtual reality, competitive advantage, management performance, KSF, KPI, risks, value, "green" marketing, UGC, WOM, FinTech, "big" Data

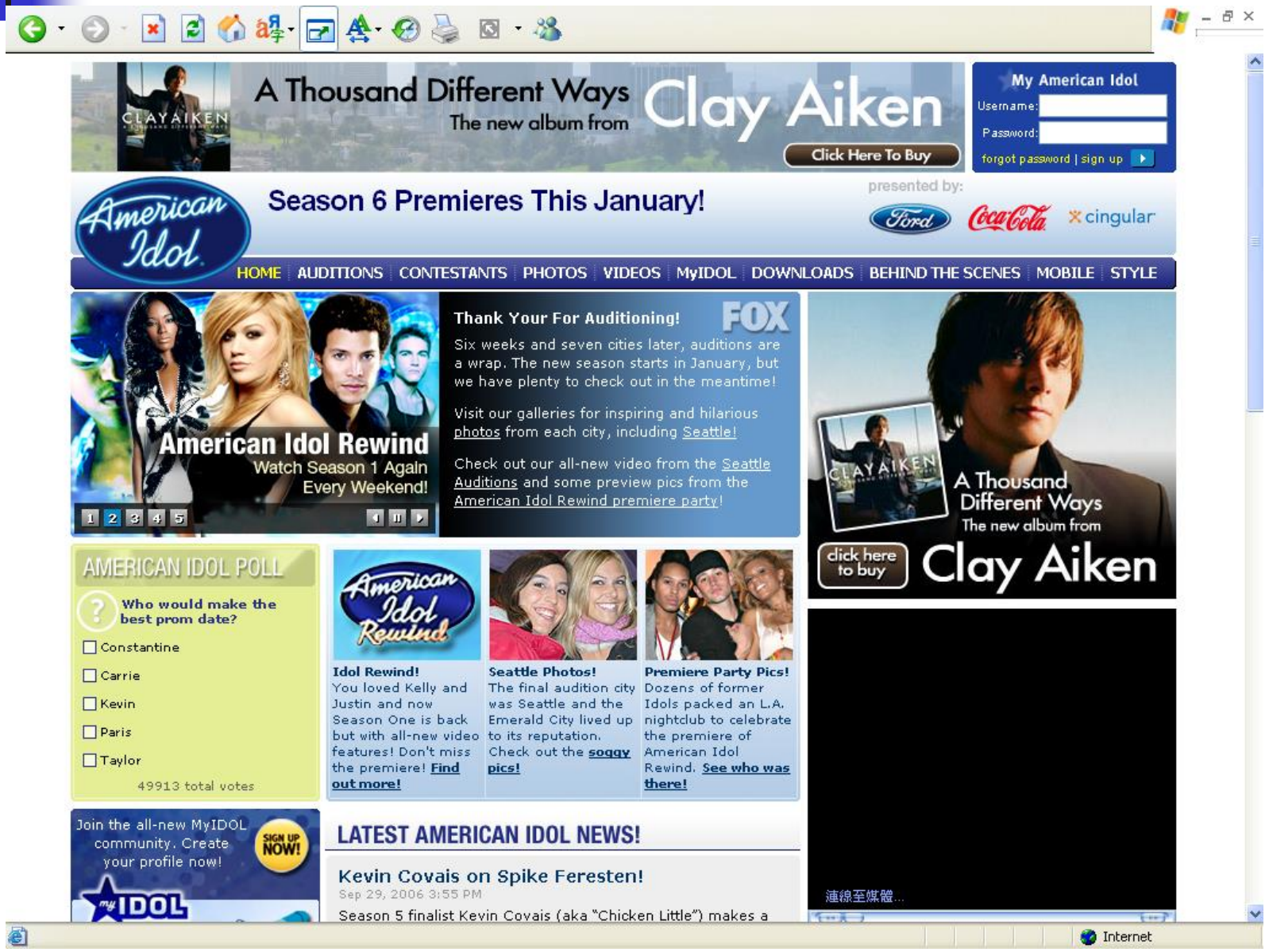
# Some are easier to measure

## Global Competitiveness Index

| | Rank (out of 133) | Score (1–7) |
|---|---|---|
| **GCI 2009–2010** | **12** | **5.2** |
| GCI 2008–2009 (out of 134) | 17 | 5.2 |
| GCI 2007–2008 (out of 131) | 14 | 5.2 |
| **Basic requirements** | **18** | **5.5** |
| 1st pillar: Institutions | 38 | 4.7 |
| 2nd pillar: Infrastructure | 16 | 5.6 |
| 3rd pillar: Macroeconomic stability | 25 | 5.3 |
| 4th pillar: Health and primary education | 15 | 6.2 |
| **Efficiency enhancers** | **17** | **5.1** |
| 5th pillar: Higher education and training | 13 | 5.4 |
| 6th pillar: Goods market efficiency | 14 | 5.1 |
| 7th pillar: Labor market efficiency | 24 | 4.8 |
| 8th pillar: Financial market sophistication | 54 | 4.4 |
| 9th pillar: Technological readiness | 18 | 5.4 |
| 10th pillar: Market size | 17 | 5.2 |
| **Innovation and sophistication factors** | **8** | **5.2** |
| 11th pillar: Business sophistication | 13 | 5.2 |
| 12th pillar: Innovation | 6 | 5.3 |



# But, some are difficult to develop



# And, many results are arguable


# Importance of the “standard”

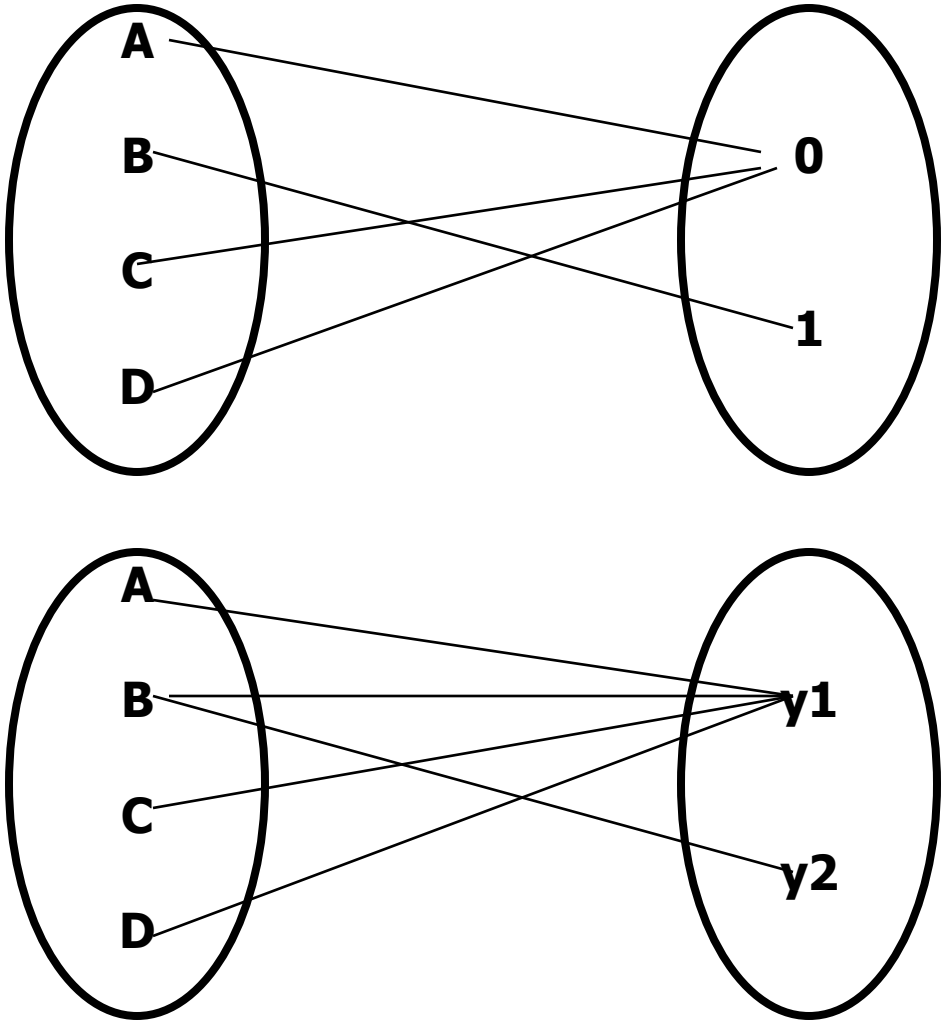
**0.5 ppm vs. 2.5ppm (*0.5*毫克／每公斤(mg／kg))**

# Measurement and Scaling

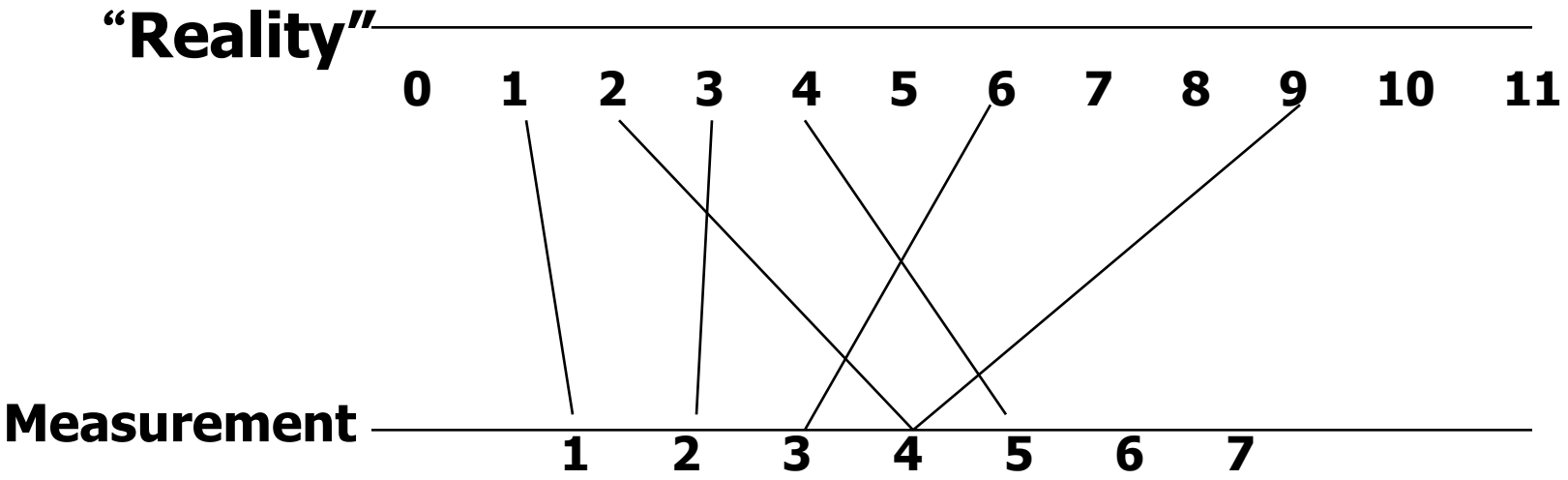
**Measurement** means assigning numbers (1, 2, 3…) or other symbols (a, b, c) to characteristics of objects according to certain pre-specified rules.

- One-to-one correspondence between the numbers and the characteristics being measured (Why is it important?)
- The rules for assigning numbers should be standardized and applied uniformly (consistent and reliable)
- Rules must not change over objects or time (reliable and valid)

# Rule of Correspondence



# Measurement and "Reality" Isomorphism

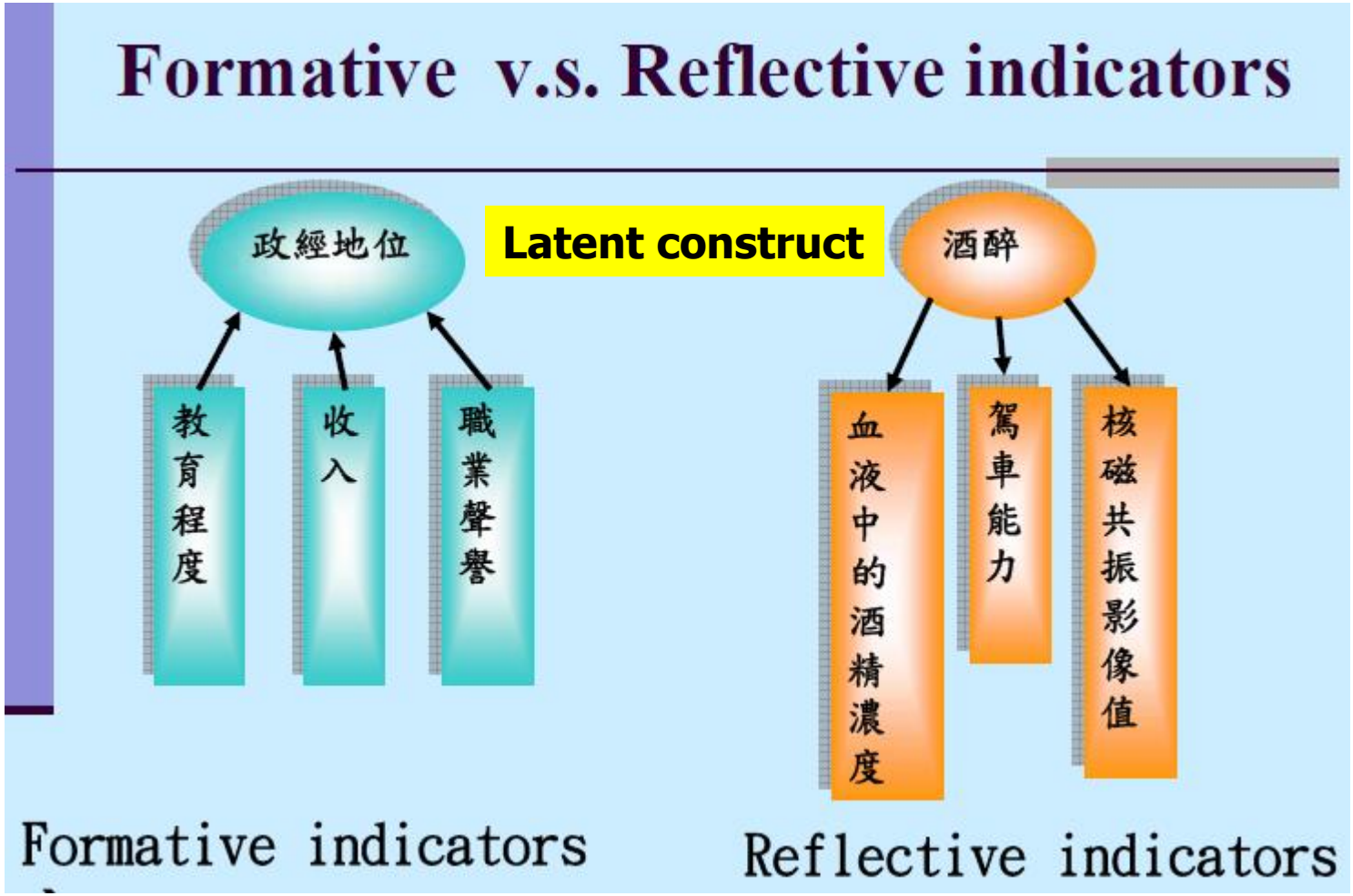



**Q1. Is the measurement tied to "reality"?**
**Q2. Do the measurement have some rational and empirical Correspondence with "reality"?**

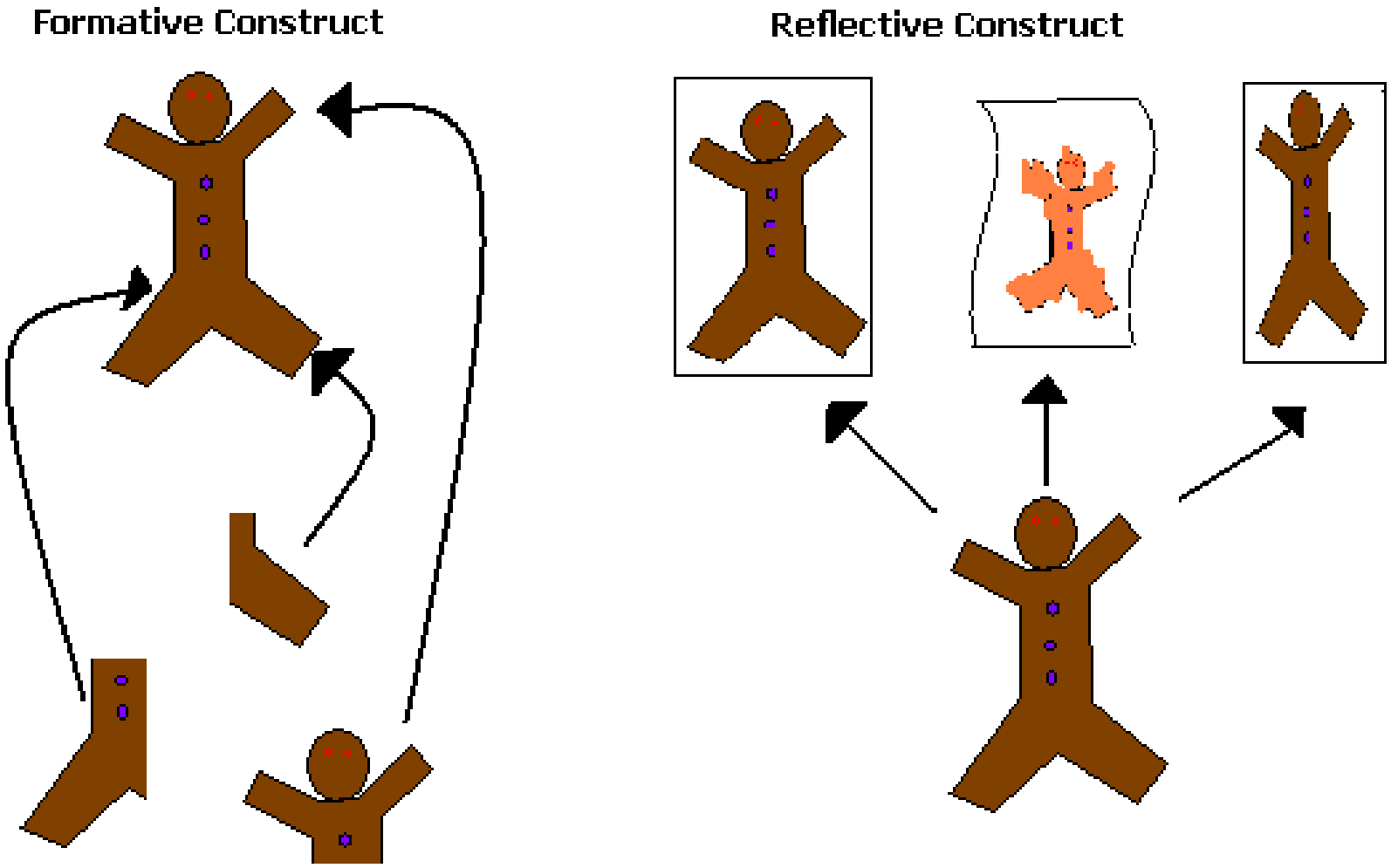
# Properties, Constructs, and Indicants of Objects

- We may say we measure the properties or the characters of the object, not the object itself (e.g., gender, hostility, loyalty, trust)
- In fact, we really measure the *indicants* of the properties of the objects (e.g., private consumption, emotional arousal, churn rate, financial ratios, CSR activities)
- **Indicants** are specified by "operational definition", which is necessary to measure a property or a construct (concept, variable, framework, hypothesis, and theory)

# Indicants: reflective vs. formative



# Formative (形成性, cause) vs. Reflective (反映性, effect)




[Graphic courtesy of Robert Sainsbury, Mississippi State University]

# Formative versus reflective indicator models

**Reflective specification model:**

$$x_i = \lambda_i \eta + \xi_i$$

**where $\lambda_i$ is the expected effect of the latent variable $\eta$ on a set of observable indicators $x_n$ and $\xi_i$ is the measurement error for the ith indicator (i = 1, 2, …. $\eta$). Here the correlation between all $x_n$ is high.**

**Formative specification model:**

$$\eta = y_1 x_1 + y_2 x_2 + \ldots + y_\eta x_\eta + \xi$$

**where $y_i$ is the expected effect of $x_i$ on the latent variable $\eta$ and is a disturbance term. Here the correlation between all $x_n$ can lie anywhere between the interval [-1;+1]**

# Reflective (反映性) Indicators

- Observed indicators (effects) are assumed to be caused by the latent variable; value changes in LV result in changing values of all reflective indicators.
- High correlations between indicators are expected and can be interpreted as high internal consistency; therefore, leaving out one specific indicator do not influence the LV content

# Formative (形成性) Indicators

- Formative indicators cause the latent variable.
- High correlations between formative indicators might occur, but are not generally expected; hence FA and Cronbachs alpha are inappropriate measure. Any elimination (addition) of formative indicators will result in reductions of scale validity.
- Formative indicators are fatally flawed and should never be used in EQS/AMOS/SEM/LISREL (Howell, Breivik and Wilcox 2007)

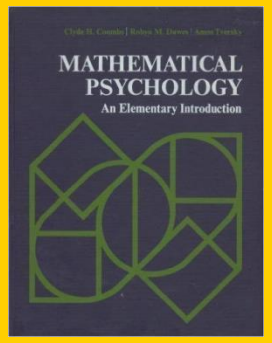
# Measurement and Scaling

- **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.

# Measurement Postulates

- (Postulates is an assumption or prerequisites to carrying out some operation.)
- Three postulates of measurement (Coombs 1953):
  1. Either (a = b) or (a ≠ b), but not both
  2. If [(a = b) and (b = c)], then (a = c)
  3. If [(a > b) and (b > c)], then (a > c)
     1. Classification
     2. Equality in comparison
     3. Transitivity (trouble in Ψ such as love, like, a friend of, accept, preference, loyalty, trust, and relationship)



# Primary Scales of Measurement

Figure 8.1

| Scale | | |
|---|---|---|
| Nominal | Numbers Assigned to Runners |   |
| Ordinal | Rank Order of Winners |   |
| 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 (gender, Tel No., region)
- 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.

# *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.
- 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.
- Preference, IQ, personality, and attitude

# *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 (e.g., global competitiveness index)
- 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 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.
- E.g., weight, money, time, growth rate, market share, new product diffusion rate, price elasticity, customer retention rate,

# Primary Scales of Measurement

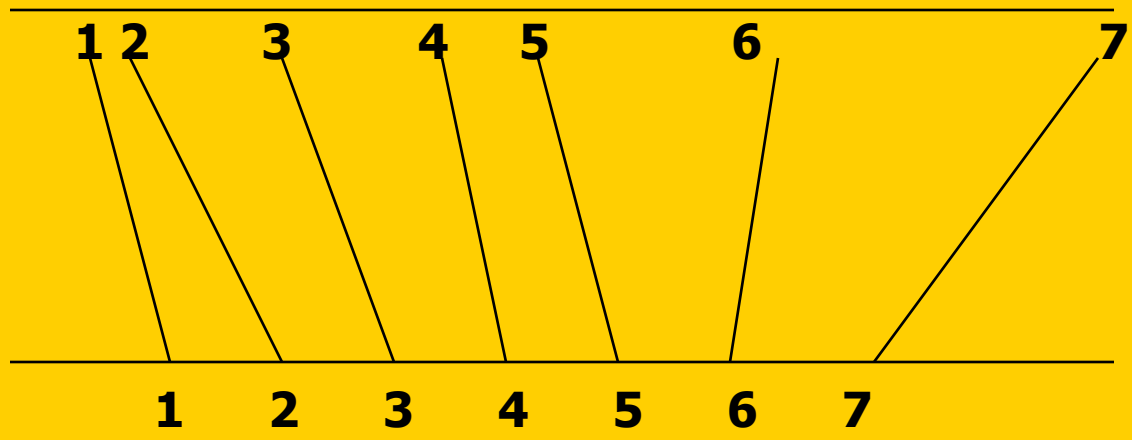
Table 8.1

| 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 |

# Beware the "equal interval" assumption (in *ordinal* and *interval* scales)

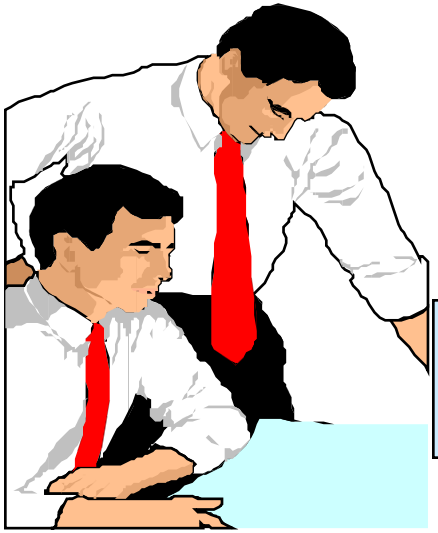
**"True" scale**

**Ordinal Scale**



# 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. (MDS and conjoint analysis)

- 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. (Likert scale)

# 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.

# *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

Perceived Value

- Under the assumption of **transitivity**, it is possible to convert paired comparison data to a rank order.

# 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.

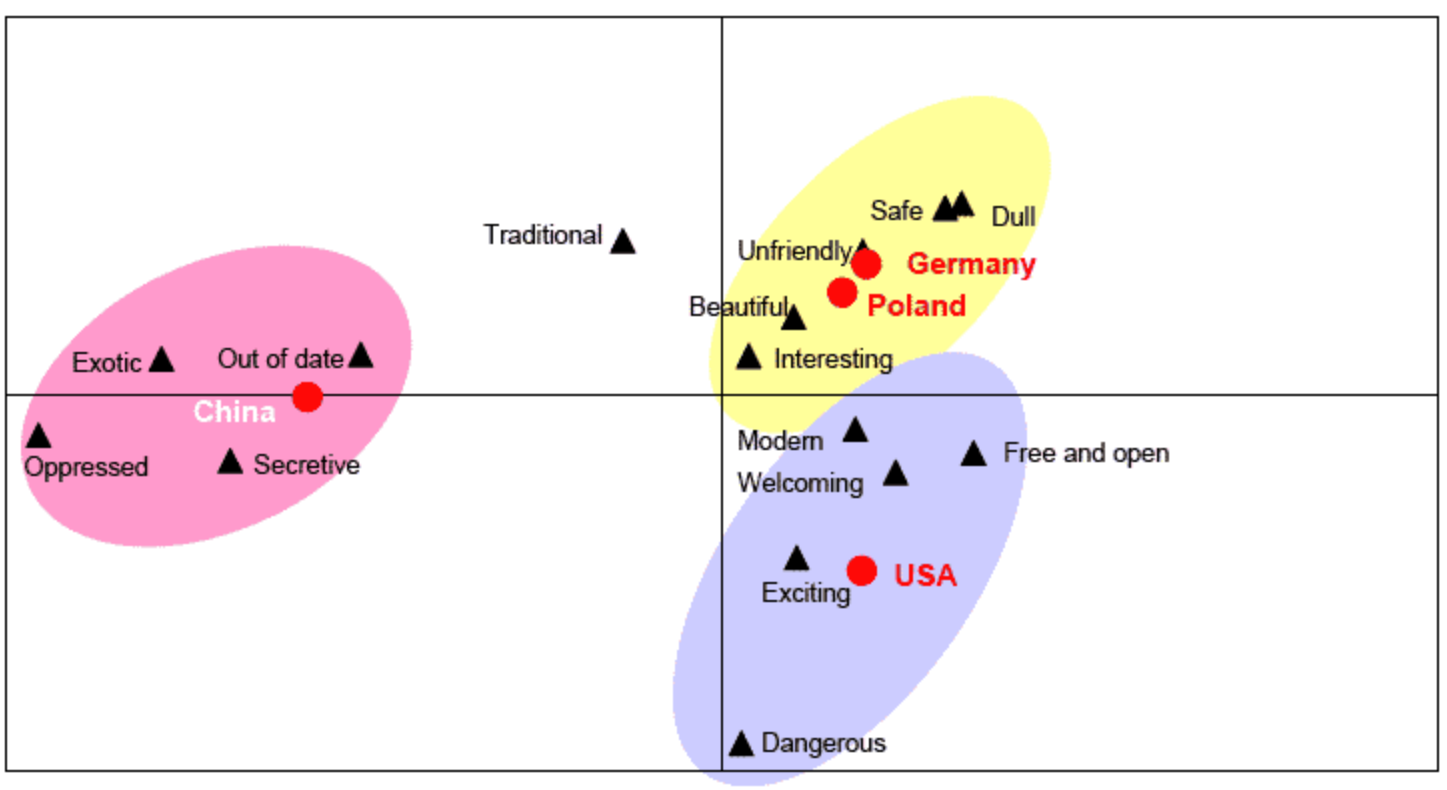
**Recording Form:**

| | 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 Comparisons (MDS): PERCEPTIONS OF CHINA IN SCOTLAND

*Q Which of these words, if any, would you use to describe [country]?*




**The Scottish Government,**
**http://www.scotland.gov.uk/Publications/2007/09/13153352/9**

# *Comparative Scaling Techniques*
# Rank Order Scaling

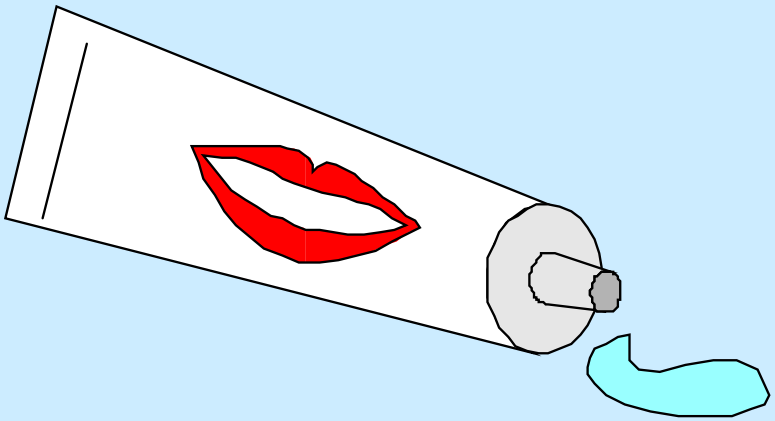
- Respondents are presented with several objects simultaneously and asked to order or rank them according to some criterion.
- It is possible that the respondent may dislike the brand ranked 1 in an absolute sense.
- Furthermore, rank order scaling also results in ordinal data.
- Only (n - 1) scaling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

# Preference for Toothpaste Brands Using Rank Order Scaling

Figure 8.4 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 Scaling Techniques*
# Constant Sum Scaling

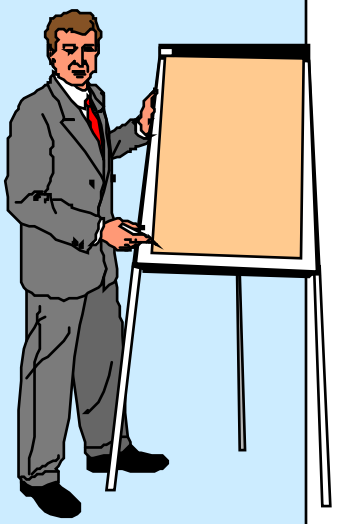
- Respondents allocate a constant sum of units, such as 100 points to attributes of a product to reflect their importance.
- If an attribute is unimportant, the respondent assigns it zero points.
- If an attribute is twice as important as some other attribute, it receives twice as many points.
- The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

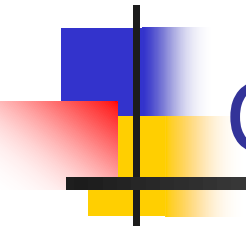


# Importance of Bathing Soap Attributes Using a Constant Sum Scale

Figure 8.5 cont.

**Form**

| Attribute | Average Responses of Three Segments | | |
|---|---|---|---|
| | 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 |



# Chapter Nine

## Measurement and Scaling:

## Noncomparative Scaling Techniques

# Noncomparative Scaling Techniques

- Respondents evaluate only one object (variable, concept, construct) at a time, and for this reason noncomparative scales are often referred to as **monadicv scales**.
- Noncomparative 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

# 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 |

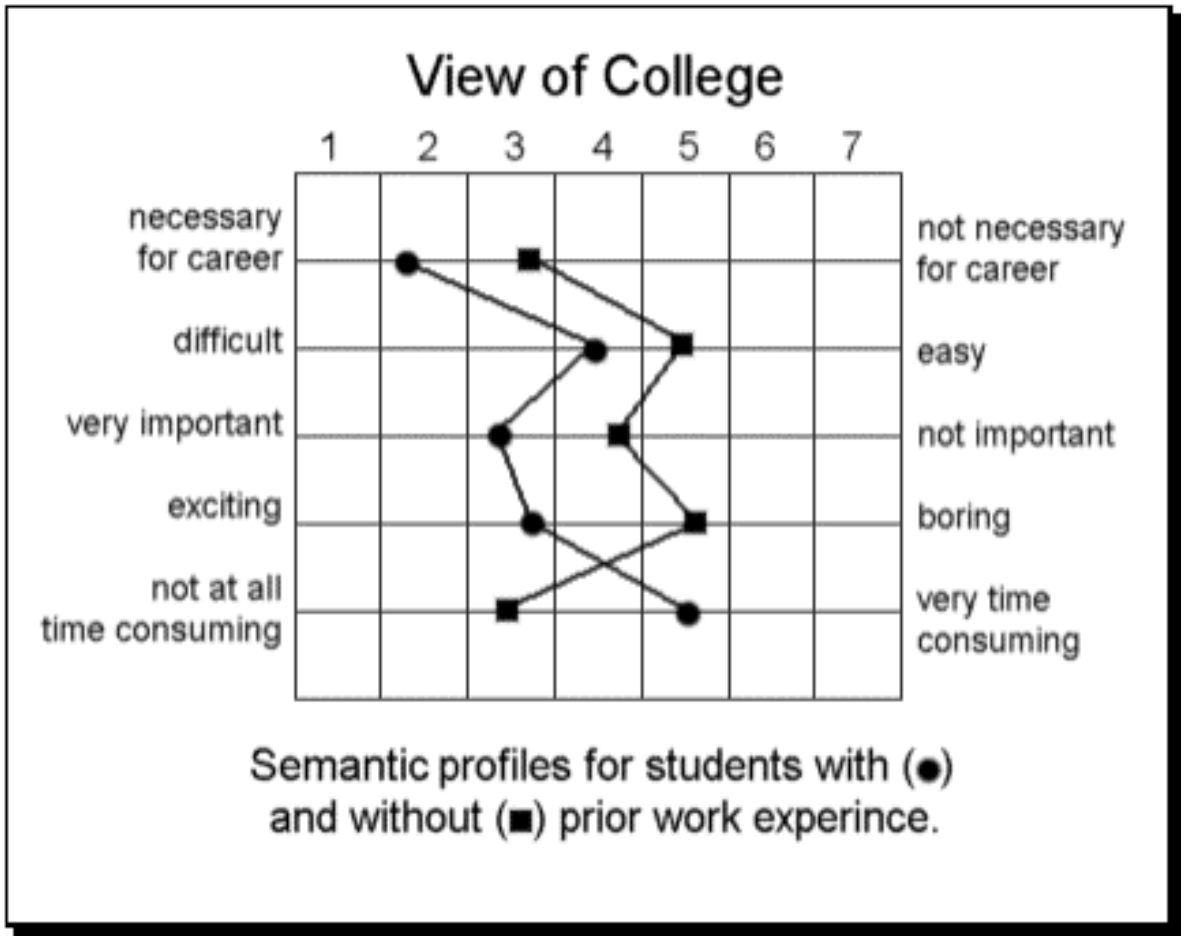
- The negative adjective or phrase sometimes appears at the left side of the scale and sometimes at the right.
- This controls the tendency of some respondents, particularly those with very positive or very negative attitudes, to mark the right- or left-hand sides without reading the labels.
- Individual 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 |
| 11) Formal | :---:---:---:---:---:---:---: | Informal |
| 12) Orthodox | :---:---:---:---:---:---:---: | Liberal |
| 13) Complex | :---:---:---:---:---:---:---: | Simple |
| 14) Colorless | :---:---:---:---:---:---:---: | Colorful |
| 15) Modest | :---:---:---:---:---:---:---: | Vain |

# Semantic Differential Scale

蛇行圖



# 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** |

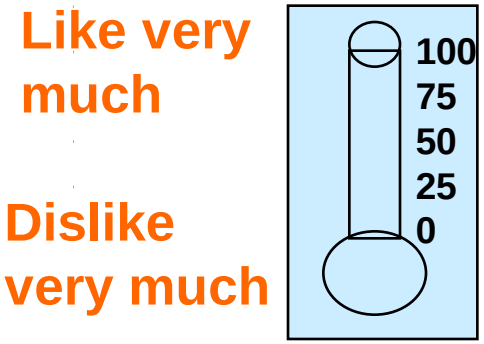
# Some Unique Rating Scale Configurations

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:



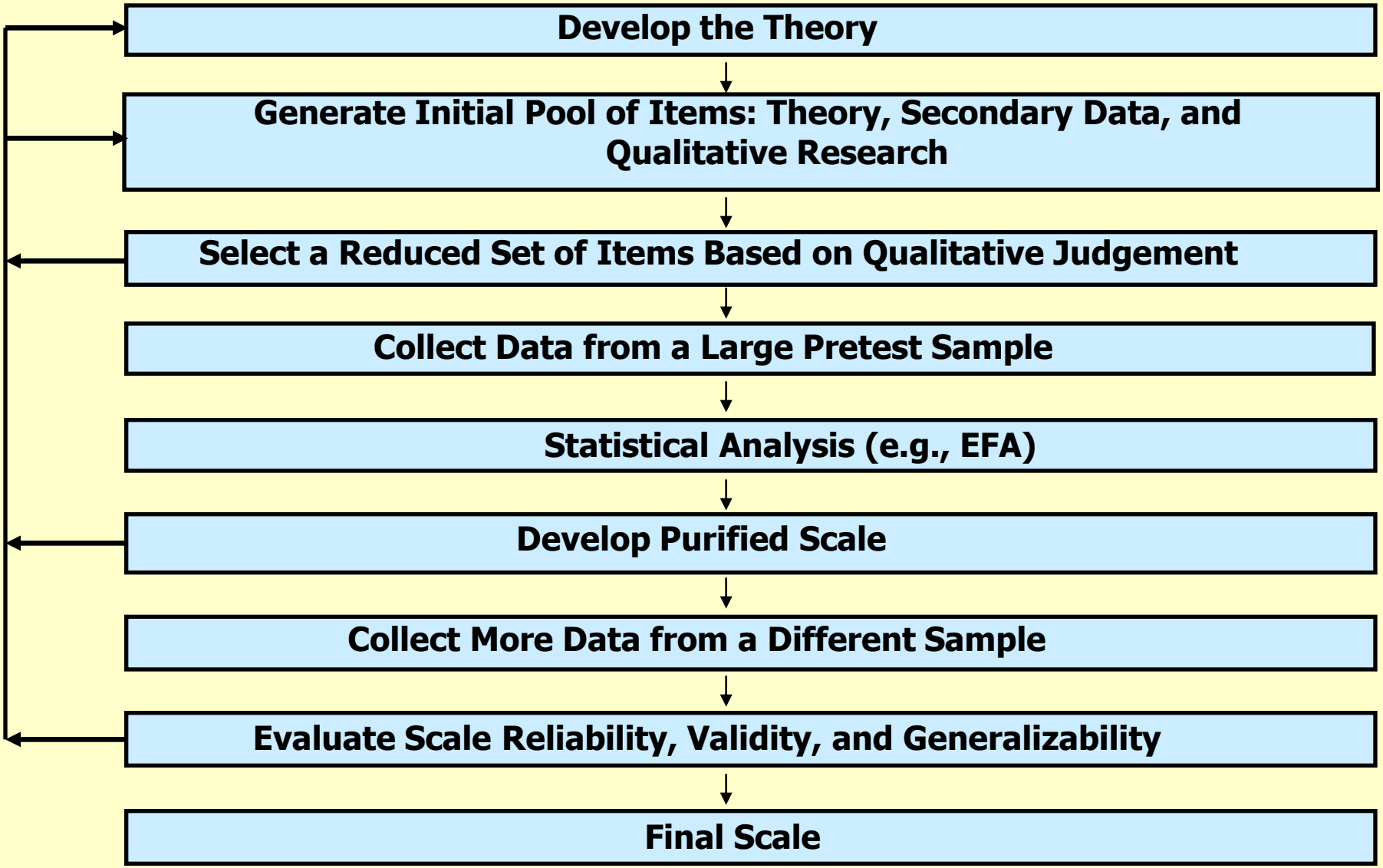

Smiling Face Scale

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.

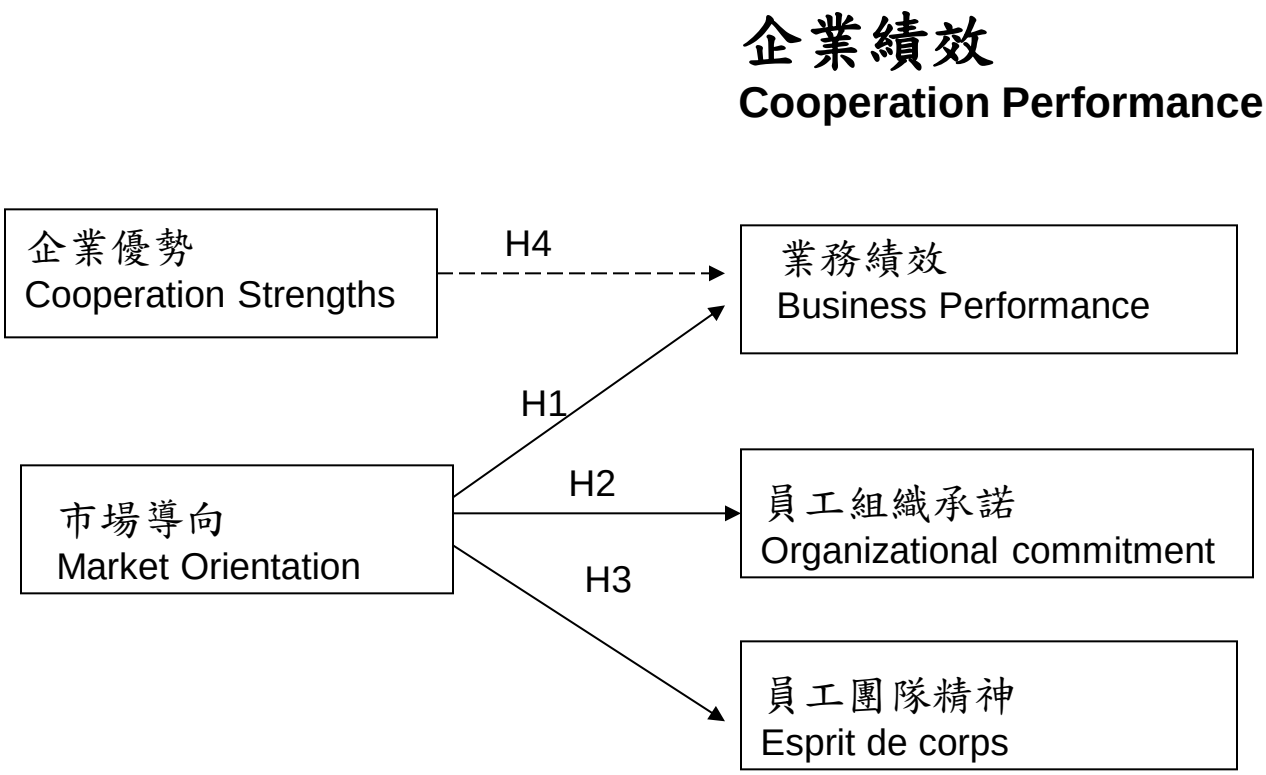
Form:

# Development of a Multi-item Scale

Figure 9.4

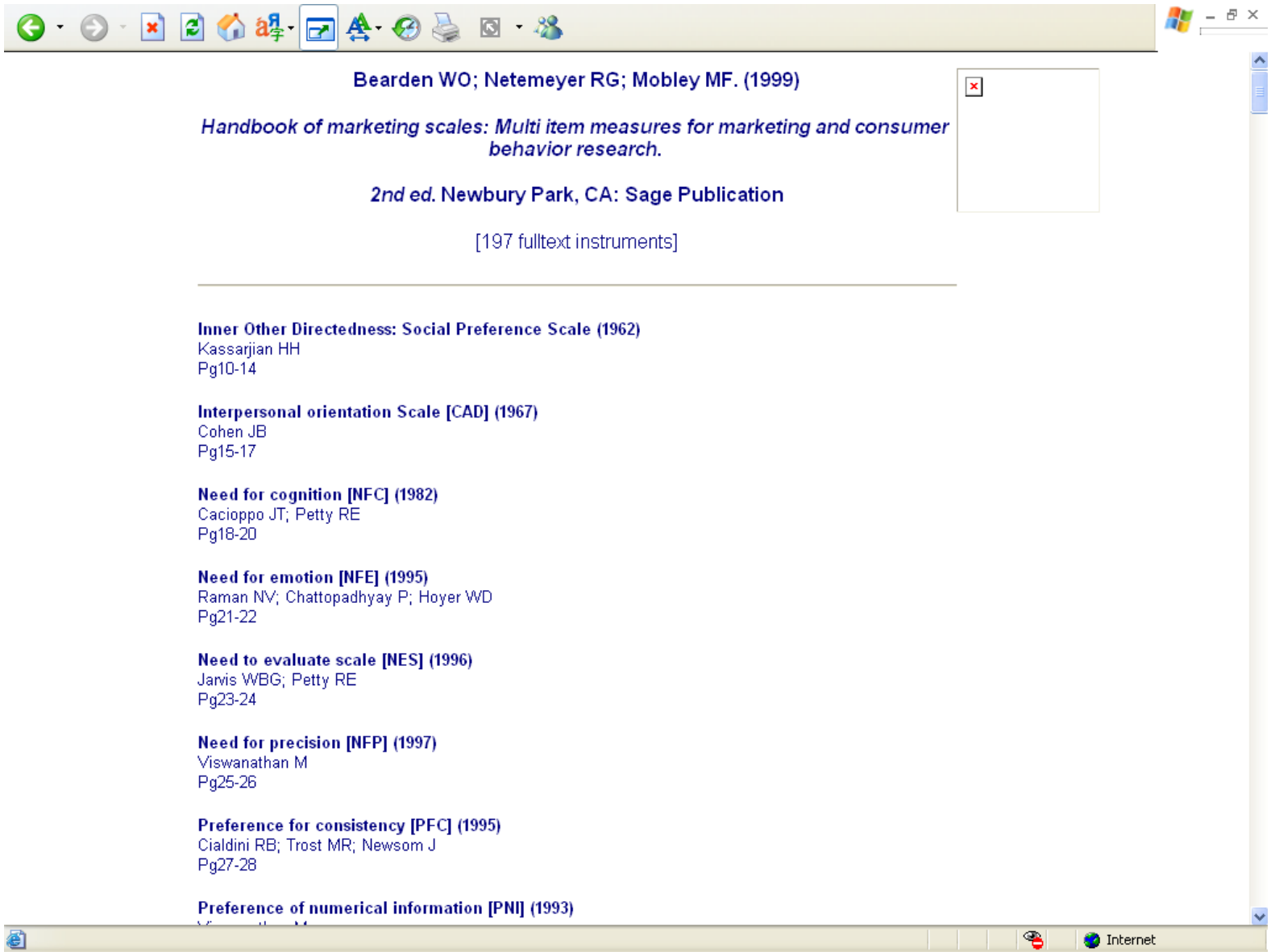


# 市場導向 研究架構




（陳佳創 2005）

# Example in marketing and OB



Bearden WO; Netemeyer RG; Mobley MF. (1999)

*Handbook of marketing scales: Multi item measures for marketing and consumer behavior research.*

*2nd ed.* Newbury Park, CA: Sage Publication

[197 fulltext instruments]

---

**Inner Other Directedness: Social Preference Scale (1962)**
Kassarjian HH
Pg10-14

**Interpersonal orientation Scale [CAD] (1967)**
Cohen JB
Pg15-17

**Need for cognition [NFC] (1982)**
Cacioppo JT; Petty RE
Pg18-20

**Need for emotion [NFE] (1995)**
Raman NV; Chattopadhyay P; Hoyer WD
Pg21-22

**Need to evaluate scale [NES] (1996)**
Jarvis WBG; Petty RE
Pg23-24

**Need for precision [NFP] (1997)**
Viswanathan M
Pg25-26

**Preference for consistency [PFC] (1995)**
Cialdini RB; Trost MR; Newsom J
Pg27-28

**Preference of numerical information [PNI] (1993)**

**Price JL. (1997).**

**Handbook of organizational measurement.**

**International Journal of Manpower.**

**V.18, No.4/5/6, pg.305-558**

---

**Absenteeism.** (1995).
Kim SW; Cyphert ST; Price JL.
Pg.314-315,319-323

**Absenteeism.** (1982).
Chadwick-Jones JK; Nicholson N; Brown C.
Pg.314-319,323

**Administrative intensity. National Organizations Study [NOS] , General Social Survey [GSS], National Opinion Research Center [NORC].** (1996).
Kalleberg AL; Knoke D; Marsden PV; Spaeth JL.
Pg.324-326, 331-334

**Administrative intensity.** (1987).
McKinley W.
Pg.324-326,329-331, 334

**Administrative intensity.** (1973).
Blau PM.
Pg.324-329, 334

**Commitment, National Organizations Study [NOS].** (1996).
Kalleberg AL; Knoke D; Marsden PV; Spaeth JL.
Pg.335-336, 339-342,347-348

**Commitment.** (1996).
Ko JW.
Pg.335-336, 342-345,347-348

**Commitment.** (1993).

Pg.388-392

**Ideology: A suggested scale.** (1997).
Price JL.
Pg.393-395

**Innovation.** (1977).
Moch MK; Morse EV.
Pg.396-397,399-403

**Innovation.** (1977).
Lewis Beck M.
Pg.396-399,401-403

**Internal labour market.** (1994).
Iverson RD; Roy P.
Pg.404-405,407-410

**Internal labour market, National Organizations Study [NOS].** (1996).
Kalleberg AL; Knoke D; Marsden PV; Spaeth JL.
Pg.404-407,410

**Work involvement, intrinsic job motivation.** (1979).
Warr P; Cook J; Wall T.
Pg.411-412,415-418, 423

**job involvement in the role [JIR] , job involvement in the section [JIS], Protestant work ethic [PWE], and work centrality [WC].** (1994).
Paullay IM; Alliger GM; Stone Romero EF.
Pg.411-412,418-423

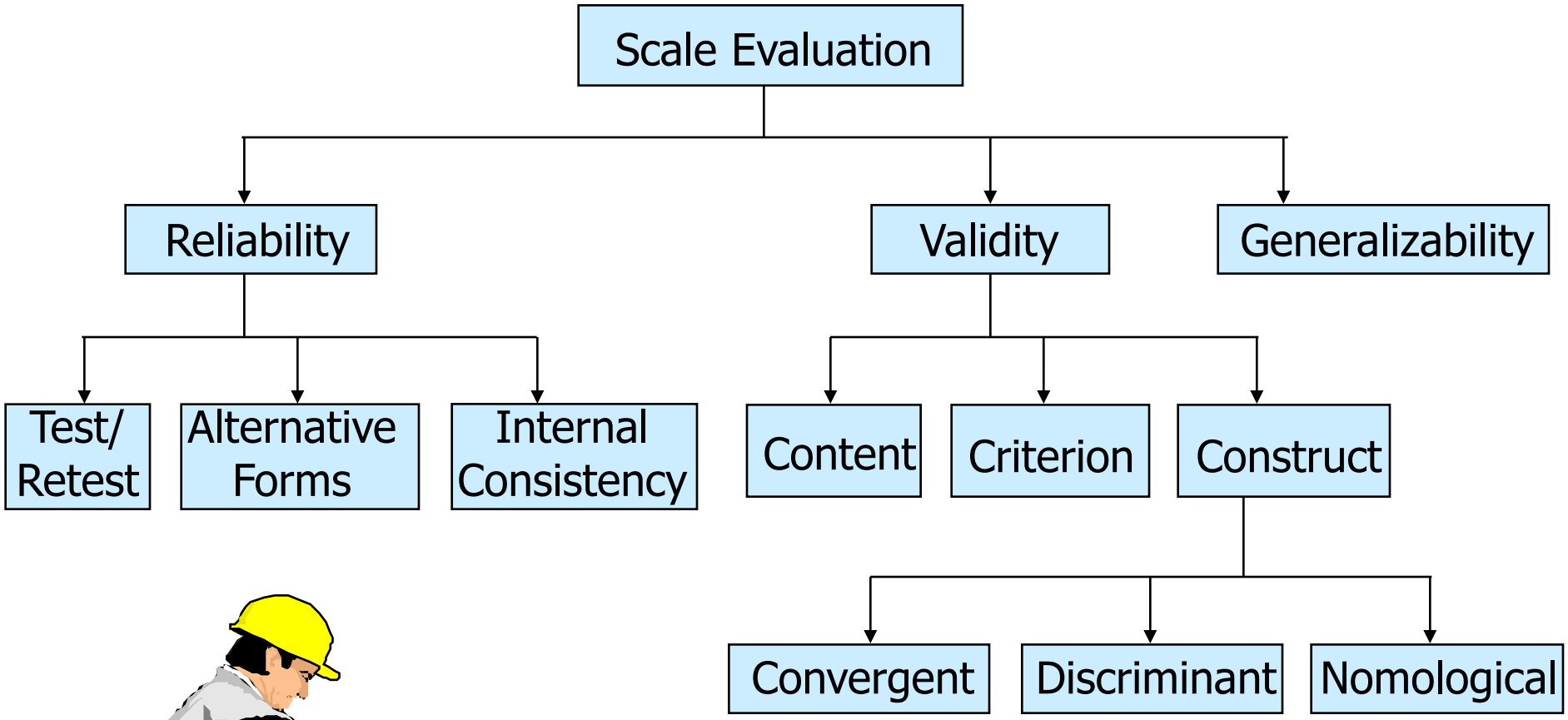
**Involvement.** (1982).
Kanungo RN.
Pg.411-415,423

**Justice.** (1996).
Kim SW; Price JL Mueller CW; Watson TW.
Pg.424-426,428

**Pay stratification: Gini index.** (1997).
Mueller CW.
Pg.429-434

# 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.

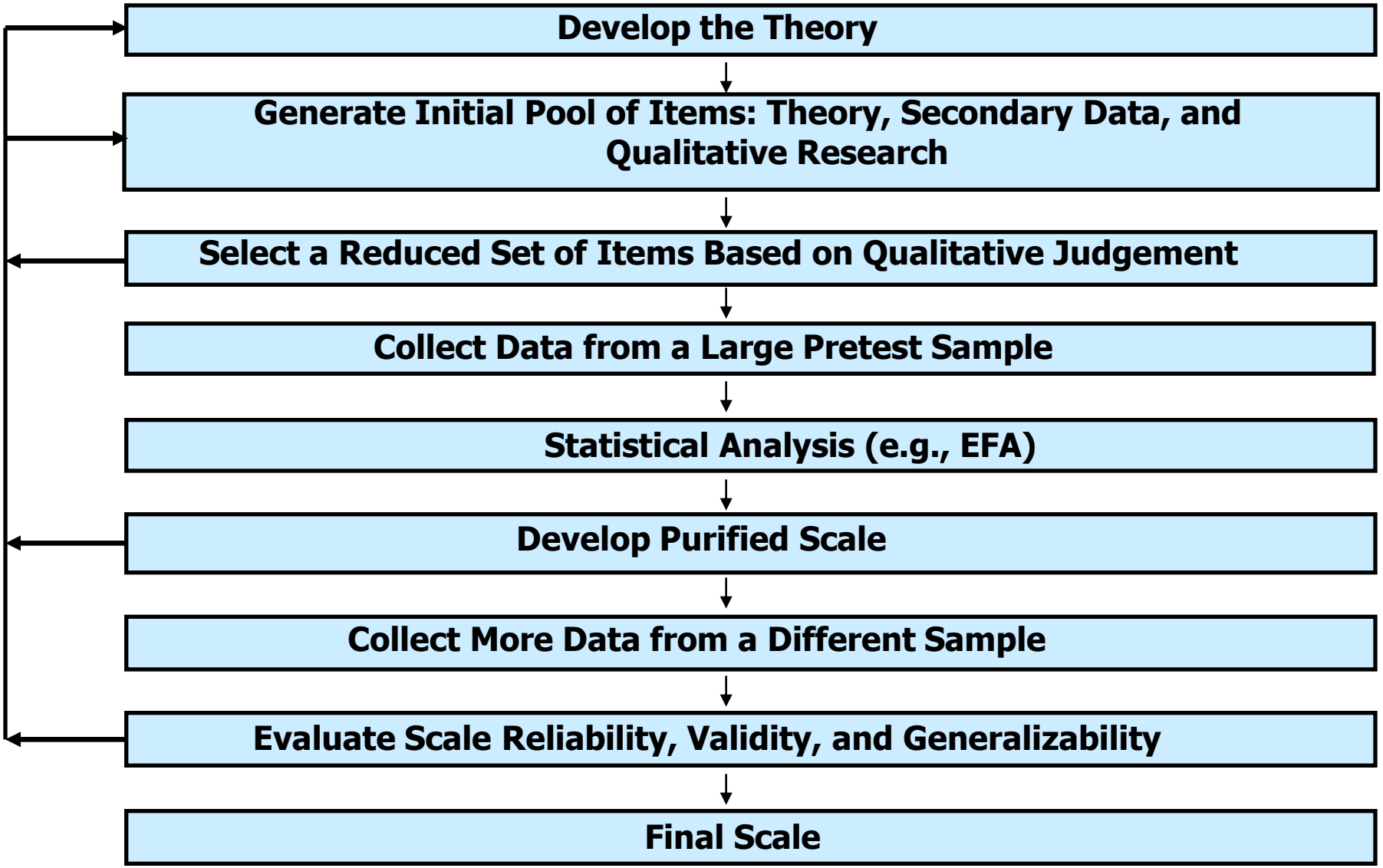
6) Mechanical factors, such as poor printing, overcrowding items in the questionnaire, and poor design.

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

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

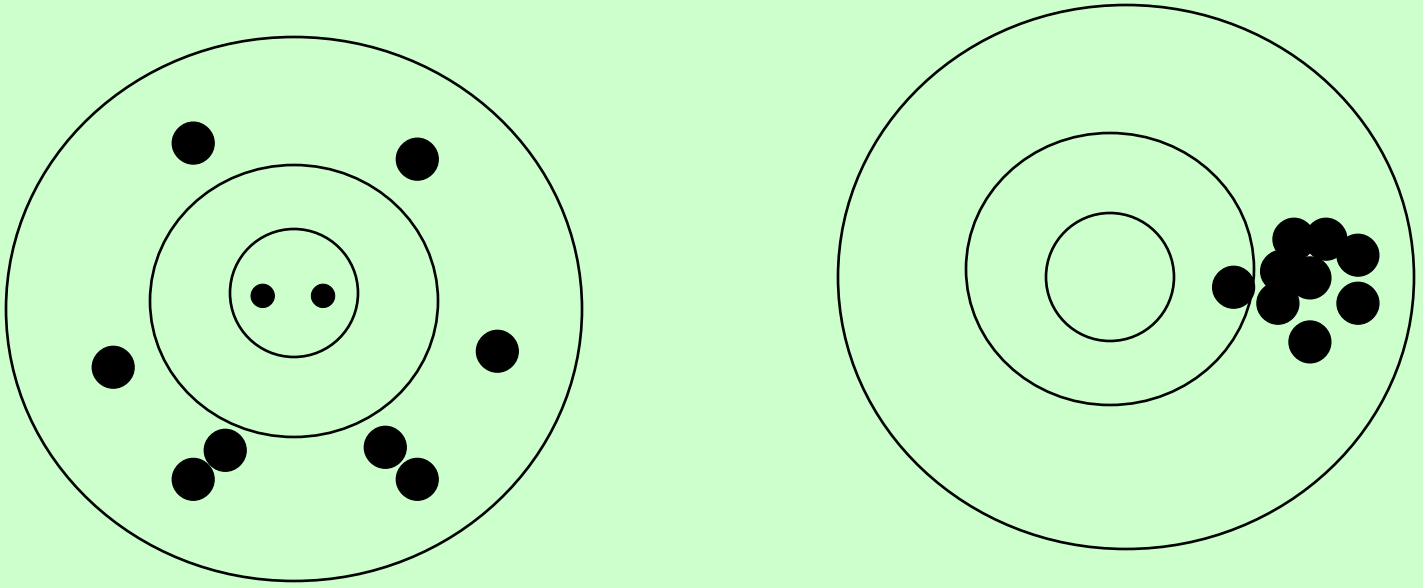
# Review: Development of a Multi-item Scale

Figure 9.4



# Reliability (信度, 可靠性)

- **Reliability:** dependability, stability, consistency, accuracy, and predictability



# Reliability

- **Reliability: dependability, stability, consistency, accuracy, and predictability**
- **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.

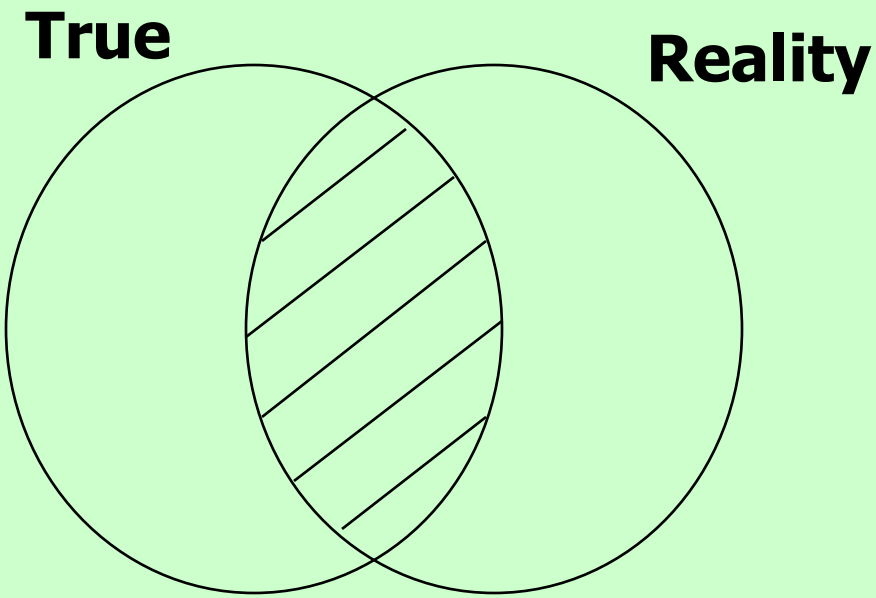
# Reliability

- **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 (效度,合法性, *Isomorphism*)

同構、類質同像

- Are we measuring what we think we are measuring?



# 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 (face) validity** is a subjective but systematic evaluation of how well the content of a scale represents the measurement task at hand (or, representative ness or sampling adequacy; judgmental)
- **Criterion validity** reflects whether a scale performs as expected in relation to other variables selected (criterion variables) as meaningful criteria (GMAT and business success, prediction)

# 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 (e.g., COO should be capable of similar interpretation in different countries)
  - **Discriminant validity** is the extent to which a measure does not correlate with other constructs from which it is supposed to differ (e.g., browsing skills of Centaur consumer are different)
  - **Nomological validity** is the extent to which the scale correlates in theoretically predicted ways with measures of different but related constructs (e.g., Centaur are field-independent and risk-prone).

# 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 (i.e., $X_R \neq 0$), 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. (e.g., I Love Taiwan because I am loyal to KMT/Ma)

| | | | |
|---|---|---|---|
| 10/31 | Questionnaire/Constructs design (10, 14) | Frequency Table | HW#0. Case 1.1 Dell Direct,p.479 (2, 4, 6) |

## Dell Running Case

Review the Dell case, Case 1.1, and questionnaire given toward the end of the book. Download the Dell case data file from the Web site for this book.

1. Recode the respondents based on total hours per week spent online into two groups: 5 hours or less (light users), and 6 hours or more (heavy users). Calculate a frequency distribution.
2. Recode the respondents based on total hours per week spent online into three groups: 5 hours or less (light users), 6 to 10 hours (medium users), and 11 hours or more (heavy users). Calculate a frequency distribution.
3. Form a new variable that denotes the total number of things that people have ever done online based on q2_1 to q2_7. Run a frequency distribution of the new variable and interpret the results. Note the missing values for q2_1 to q2_7 are coded as 0.
4. Recode q4 (overall satisfaction) into two groups: Very satisfied (rating of 1), and somewhat satisfied or dissatisfied (ratings of 2, 3, and 4). Calculate a frequency distribution of the new variable and interpret the results.
5. Recode q5 (would recommend) into two groups: Definitely would recommend (rating of 1), and probably would or less likely to recommend (ratings of 2, 3, 4, and 5). Calculate a frequency distribution of the new variable and interpret the results.
6. Recode q6 (likelihood of choosing Dell) into two groups: Definitely would choose (rating of 1), and probably would or less likely to choose (ratings of 2, 3, 4, and 5). Calculate a frequency distribution of the new variable and interpret the results.
7. Recode q9_5per into three groups: Definitely or probably would have purchased (ratings of 1 and 2), Might or might not have purchased (rating of 3), and Probably or definitely would not have purchased (ratings of 4, and 5). Calculate a frequency distribution of the new variable and interpret the results.
8. Recode q9_10per into three groups: Definitely or probably would have purchased and might or might not have purchased (ratings of 1, 2, and 3), Probably would not have purchased (rating of 4), and Definitely would not have purchased (rating of 5). Calculate a frequency distribution of the new variable and interpret the results.
9. Recode the demographics as follows. (a) Combine the two lowest education (q11) categories into a single category. Thus, Some high school or less and High school graduate will be combined into a single category labeled High school graduate or less. (b) Recode age (q12) into four new categories: 18 to 29, 30 to 39, 40 to 49, and 50 or older. (c) Combine the two lowest income (q13) categories into a single category labeled Under $30,000. Calculate frequency distributions of the new variables and interpret the results.

| | | | |
|---|---|---|---|
| 11/14 | ANOVA & Regression (16, 17) | Discriminant | HW#1. Nike (b, c, d, e, f, g, h) |

## Internet and Computer Exercises

SPSS Data File SAS Data File

1. In a pretest, data on Nike were obtained from 45 respondents. These data are given in the following table, which gives the usage, sex, awareness, attitude, preference, intention, and loyalty toward Nike of a sample of Nike users. Usage has been coded as 1, 2, or 3, representing light, medium, or heavy users. The sex has been coded as 1 for females and 2 for males. Awareness, attitude, preference, intention, and loyalty are measured on 7-point Likert-type scales (1 = very unfavorable, 7 = very favorable). Note that five respondents have missing values that are denoted by 9.

| Number | Usage | Sex | Awareness | Attitude | Preference | Intention | Loyalty |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 2 | 7 | 6 | 5 | 5 | 6 |
| 2 | 1 | 1 | 2 | 2 | 4 | 6 | 5 |
| 3 | 1 | 1 | 3 | 3 | 6 | 7 | 6 |
| 4 | 3 | 2 | 6 | 5 | 5 | 3 | 2 |
| 5 | 3 | 2 | 5 | 4 | 7 | 4 | 3 |
| 6 | 2 | 2 | 4 | 3 | 5 | 2 | 3 |
| 7 | 2 | 1 | 5 | 4 | 4 | 3 | 2 |
| 8 | 1 | 1 | 2 | 1 | 3 | 4 | 5 |
| 9 | 2 | 2 | 4 | 4 | 3 | 6 | 5 |
| 10 | 1 | 1 | 3 | 1 | 2 | 4 | 5 |
| 11 | 3 | 2 | 6 | 7 | 6 | 4 | 5 |
| 12 | 3 | 2 | 6 | 5 | 6 | 4 | 4 |
| 13 | 1 | 1 | 4 | 3 | 3 | 1 | 1 |
| 14 | 3 | 2 | 6 | 4 | 5 | 3 | 2 |
| 15 | 1 | 2 | 4 | 3 | 4 | 5 | 6 |
| 16 | 1 | 2 | 3 | 4 | 2 | 4 | 2 |
| 17 | 3 | 1 | 7 | 6 | 4 | 5 | 3 |
| 18 | 2 | 1 | 6 | 5 | 4 | 3 | 2 |
| 19 | 1 | 1 | 1 | 1 | 3 | 4 | 5 |
| 20 | 3 | 1 | 5 | 7 | 4 | 1 | 2 |
| 21 | 3 | 2 | 6 | 6 | 7 | 7 | 5 |
| 22 | 2 | 2 | 2 | 3 | 1 | 4 | 2 |
| 23 | 1 | 1 | 1 | 1 | 3 | 2 | 2 |
| 24 | 3 | 1 | 6 | 7 | 6 | 7 | 6 |
| 25 | 1 | 2 | 3 | 2 | 2 | 1 | 1 |

Analyze the Nike data to answer the following questions. In each case, formulate the null and the alternative hypotheses and conduct the appropriate statistical test(s).

a. Obtain a frequency distribution for each of the following variables and calculate the relevant statistics: awareness, attitude, preference, intention, and loyalty toward Nike.
b. Conduct a cross-tabulation of the usage with sex. Interpret the results.
c. Does the awareness for Nike exceed 3.0?
d. Do the males and females differ in their awareness for Nike? Their attitude toward Nike? Their loyalty for Nike?
e. Do the respondents in the pretest have a higher level of awareness than loyalty?
f. Does awareness of Nike follow a normal distribution?
g. Is the distribution of preference for Nike normal?
h. Assume that awareness toward Nike was measured on an ordinal scale rather than an interval scale. Do males and females differ in their awareness toward Nike?
i. Assume that loyalty toward Nike was measured on an ordinal scale rather than an interval scale. Do males and females differ in their loyalty toward Nike?
j. Assume that attitude and loyalty toward Nike were measured on an ordinal scale rather than an interval scale. Do the respondents have greater awareness of Nike than loyalty for Nike?