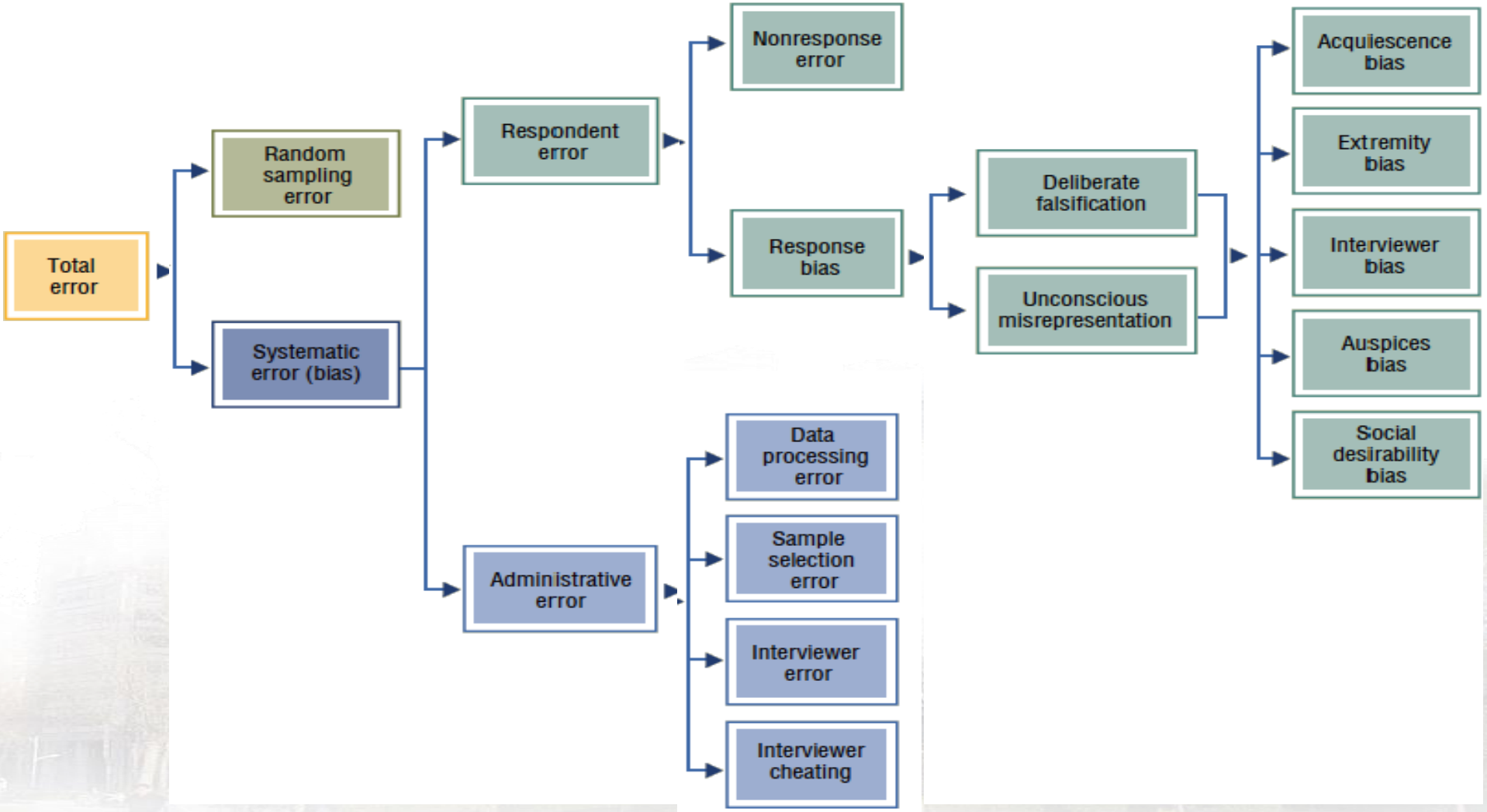
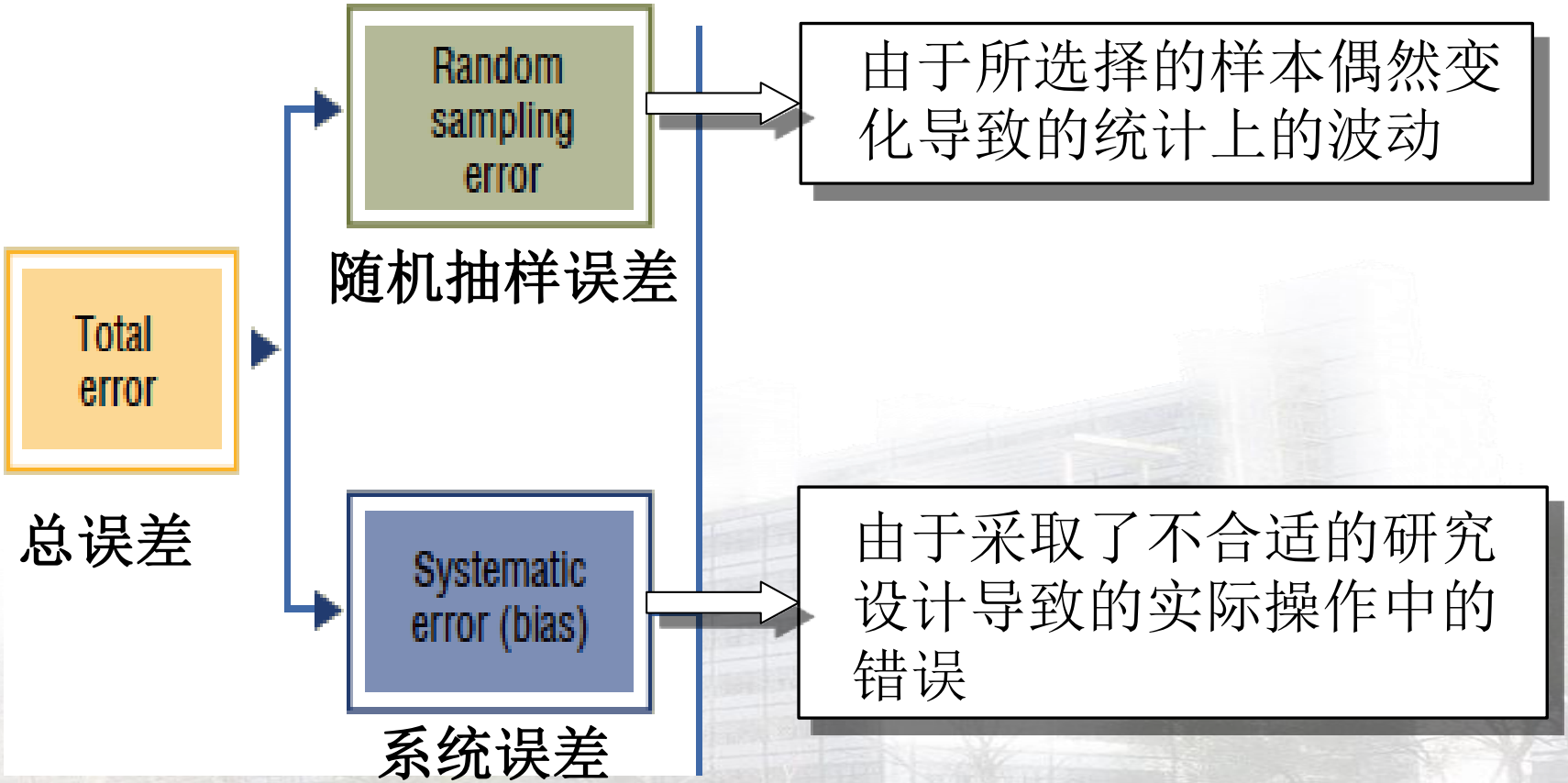
偏误和效度
(Bias and Validity)

Survey Errors

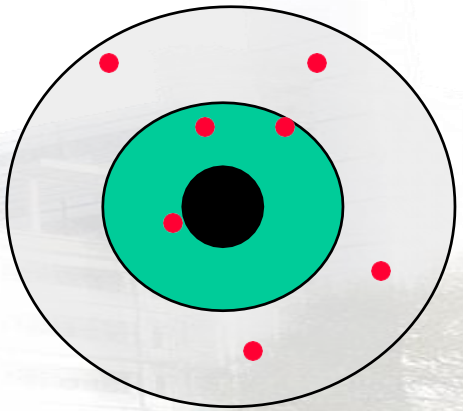


Tree Diagram of Total Survey Error



随机抽样误差 Random Sampling Error

- 样本统计量的**固有变异性**的大小，是样本统计量和总体参数之间的差距，是在选取样本时因机遇造成的。



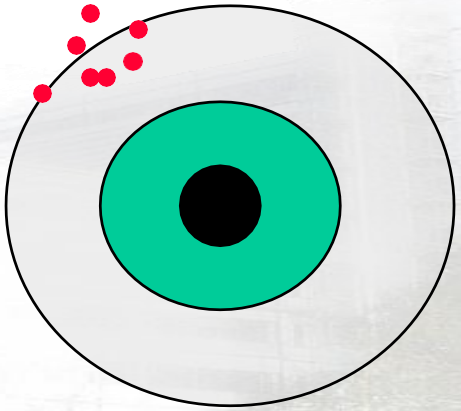
7am	9am	1pm	3pm	6pm	9pm	12pm
123	122	124.5	121	123	122.5	121.5

系统误差 Systematic (non-sampling) Error

- 由研究设计中一些不正确的方面导致的，如样本选择的错误，样本范围错误等。不是由**偶然因素**波动导致的而是由研究者产生的。

Does not begin at zero

Precision but no validity

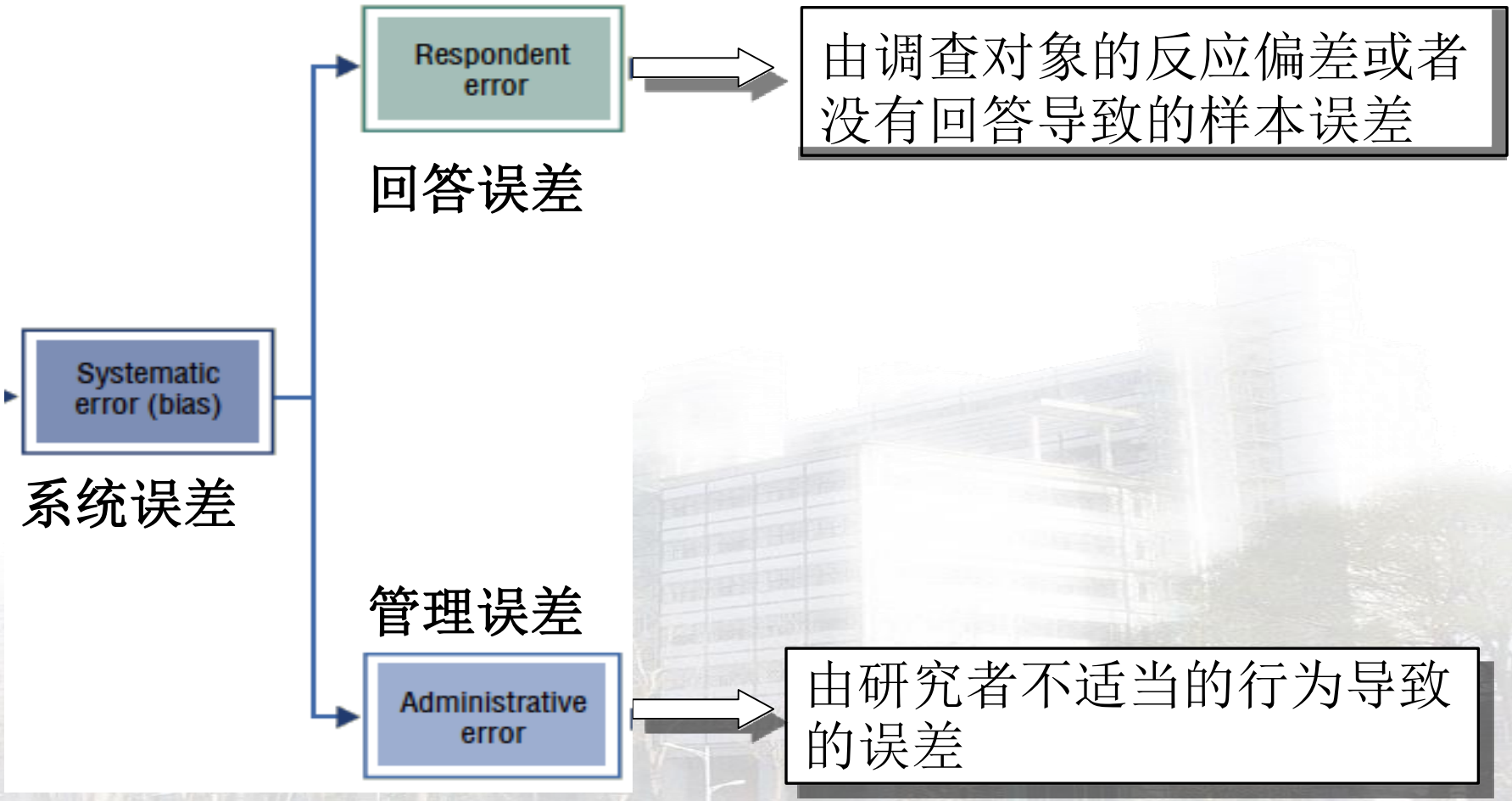


Systematic error!

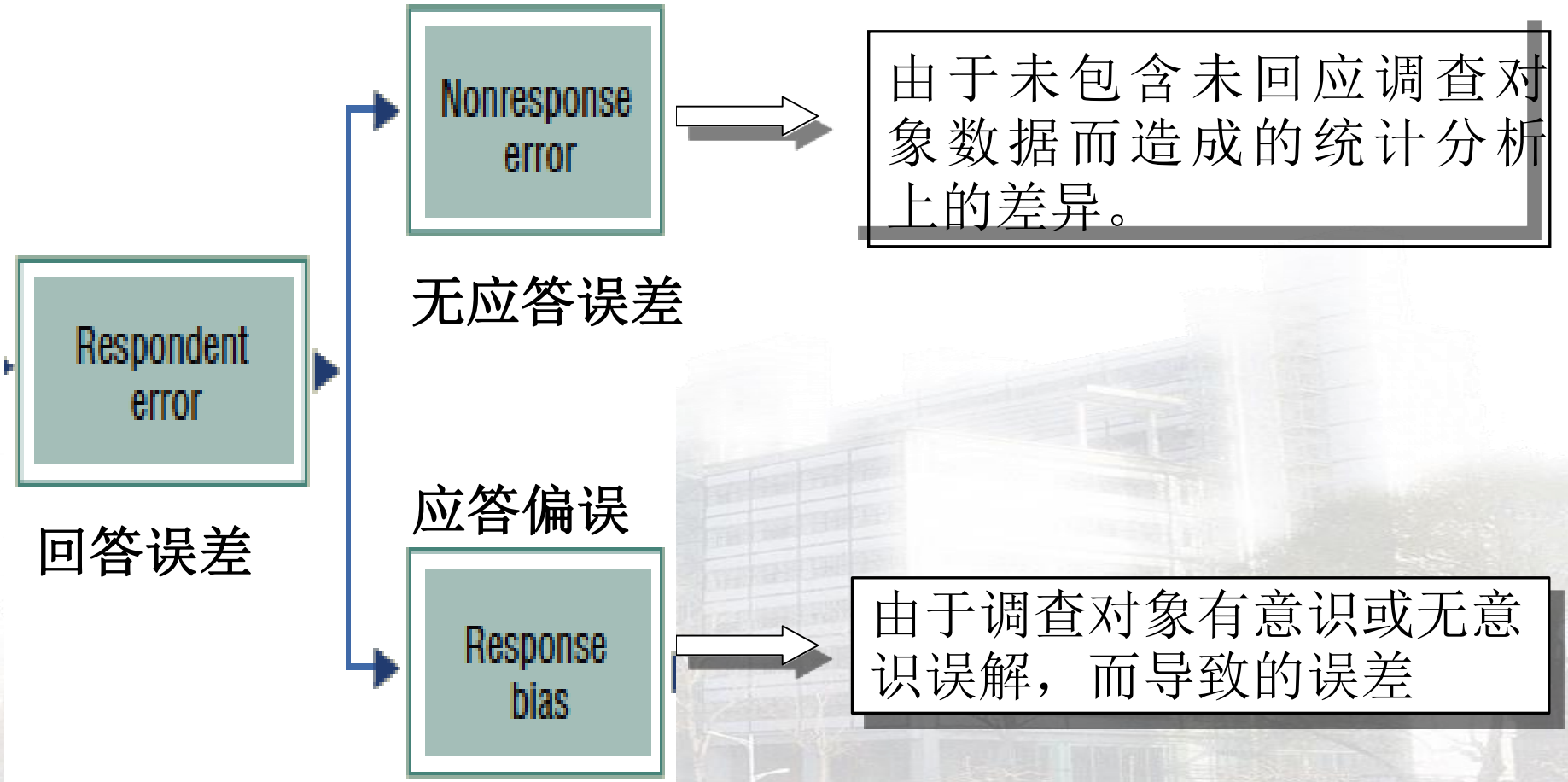
Random Sampling vs. Systematic

项目	系统误差	随机抽样误差
产生原因	固定因素，有时不存在	不定因素，总是存在
分类	方法误差、主观误差	环境的变化因素、主观的变化因素等
性质	重现性、单向性（或周期性）、可测性	服从概率统计规律、不可测性
影响	准确度	精密度
消除或减小的方法	校正	增加测定的次数

Diagram of Total Survey Error



Tree Diagram of Total Survey Error



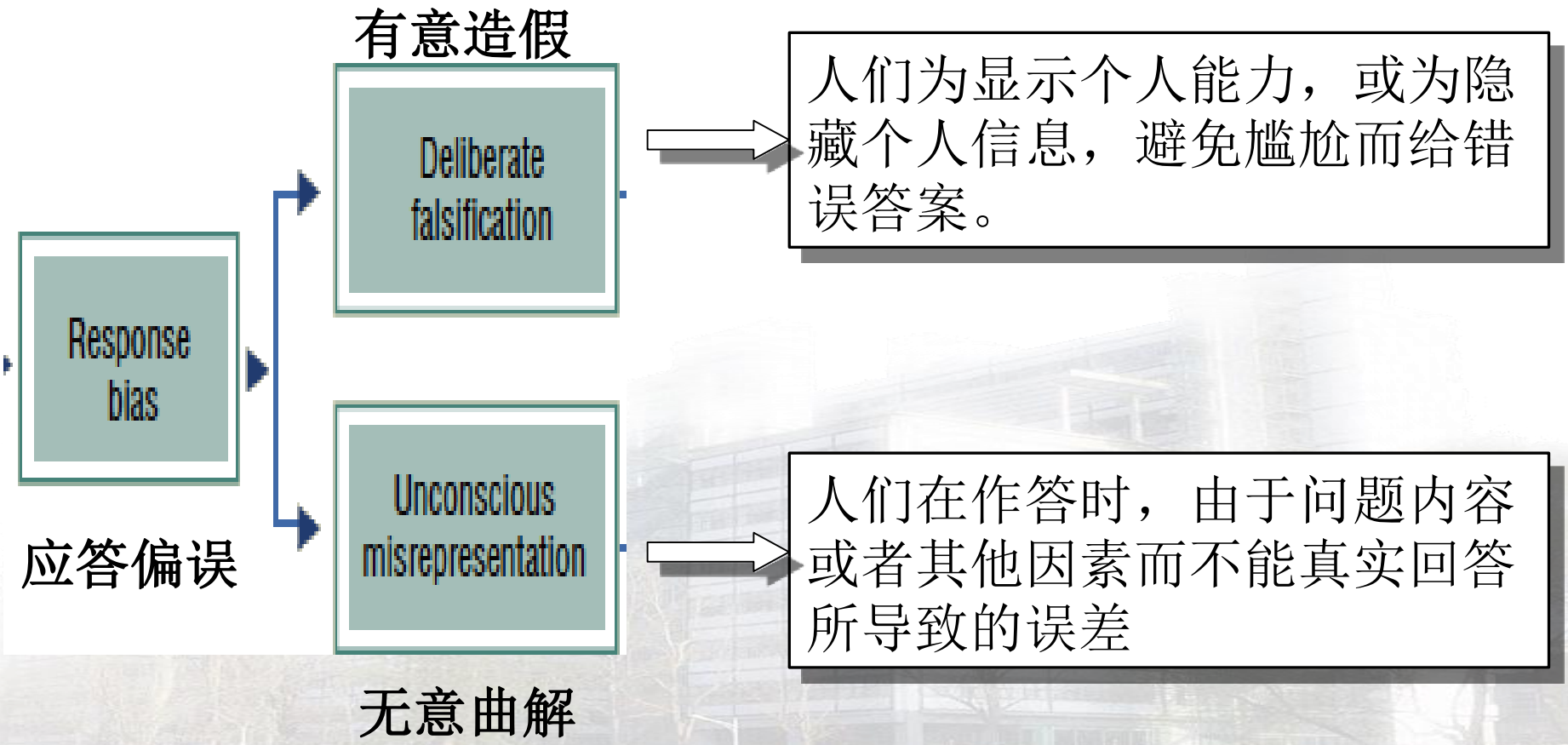
Respondent error -- Non response error

无应答误差就是实际的调查对象回答结果与理论上100%的调查结果的差别

Non response error…statistical difference between actual respondents & a perfect,100% response rate

- 不回答的调查对象——那些拒绝合作回答问题而影响结果的人们
  - 1. 对工作量的感受 Perceived amount of work required
  - 2. 对问题的内在兴趣 Intrinsic interest of the topic
  - 3. 样本的特点 Characteristics of the sample
- 没有联系 No contacts/Not-at-homes

Diagram of Total Survey Error



Respondent error -- Response bias

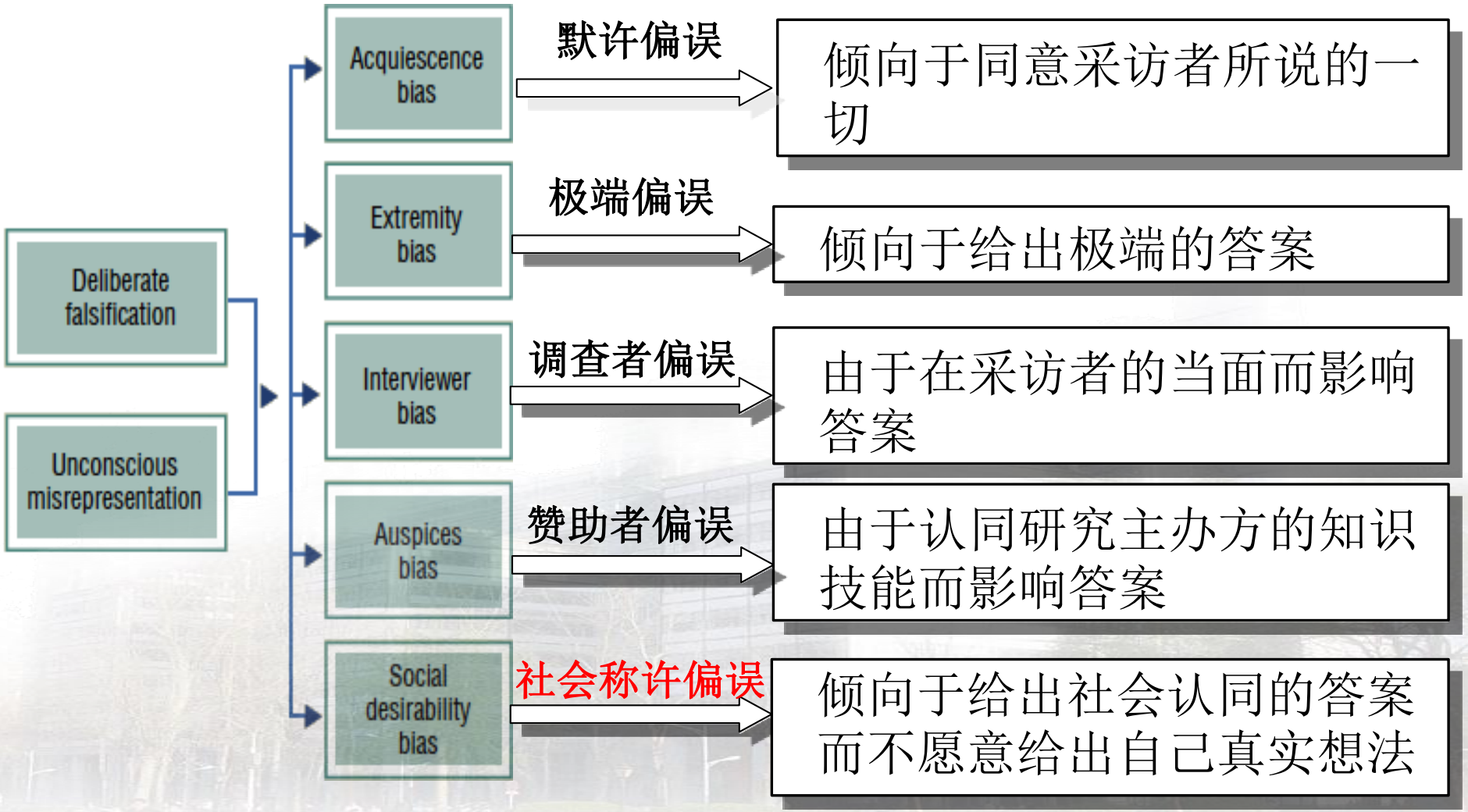
- 显示个人能力
- 隐藏个人隐私
- 避免尴尬
- 对采访者反感
- 迎合采访者

有意造假
Deliberate Falsification

无意曲解
Unconscious Misrepresentation

- 提问格式或者内容
- 访谈环境
- 误解问题
- 忘记详细说明
- 意想不到的问题
- 难以表达想法

Tree Diagram of Total Survey Error



Respondent error -- Response bias

- **默许偏误 Acquiescence Bias**
  - 一种倾向于同意所有或绝大多数问题而导致的一种反应偏误

Strongly disagree				Strongly agree
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

- **极端偏误 Extremity Bias**
  - 调查对象回答问题时倾向于给出极端的答案

Respondent error -- Response bias

- **调查者偏误 Interviewer Bias**
- 当采访者在场时而影响调查对象答案产生的偏误

A response bias when the presence of the interviewer influences answers.

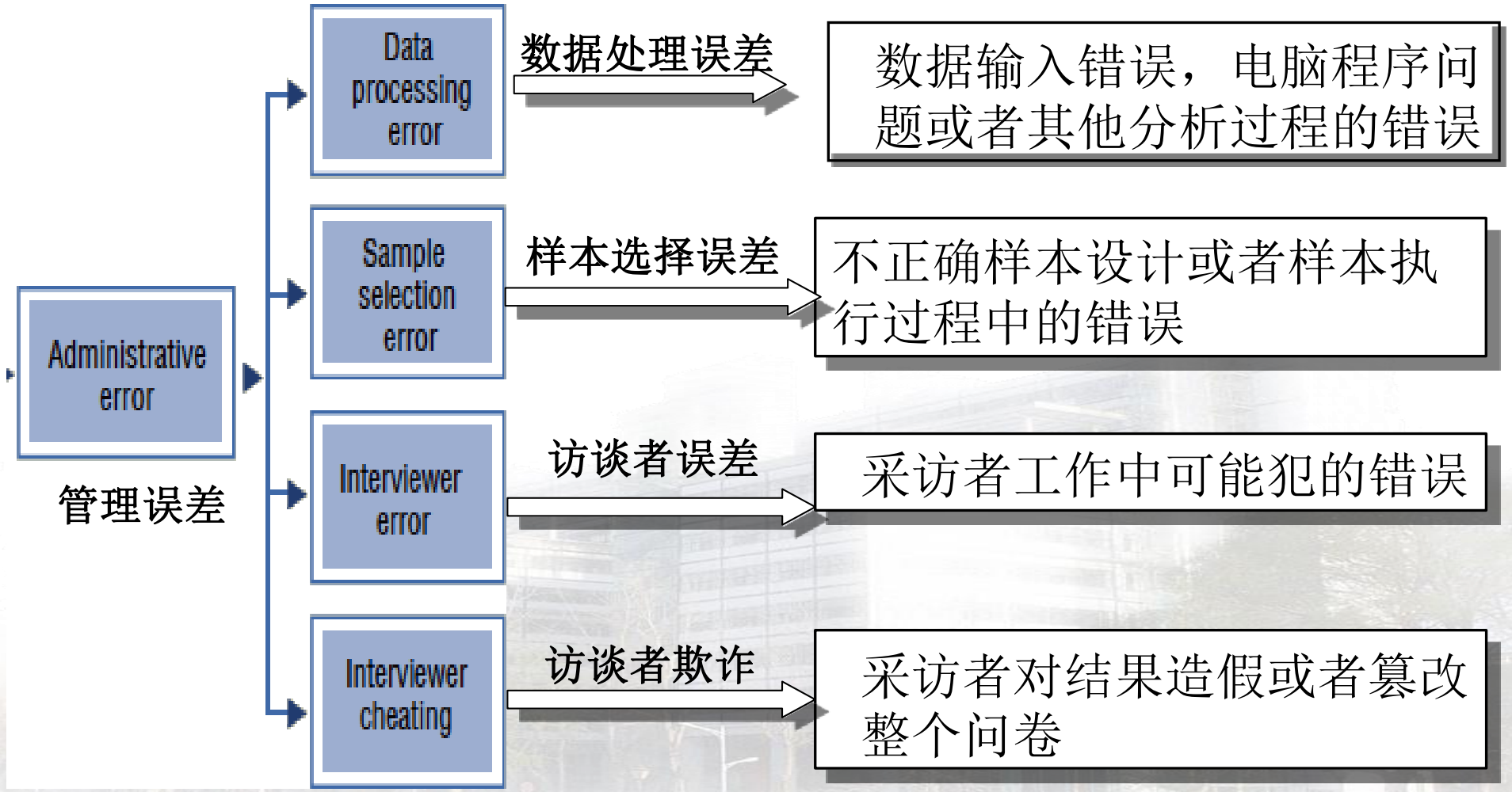
- **赞助单位偏误 Auspices Bias**
- 调查对象受到研究主办组织的影响而产生的偏误

Bias in the responses of subjects caused by the respondents being influenced by the organization conducting the study.

Respondent error -- Response bias

- **社会称许偏误 Social Desirability Bias**
- 调查对象有意识或者无意识的以社会主流意识，而不是自身真实情况反映所导致的偏误

Tree Diagram of Total Survey Error



Administrative Errors

- 研究工作中不正确的管理或者执行导致的错误原因
  - 粗心 Carelessness
  - 混淆 Confusion
  - 忽视 Neglect
  - 疏忽 Omission
  - 其他的错误 Some other blunder

信度和效度

Reliability and Validity

Measurement Error

- **观察值** Observed score

观察值(Observed Score)= 实际值(True score) + 测量误差(Measurement error)

(actual score obtained) **(stable score)** **(chance/random error)**

(systematic error)

- **实际值** True Score

完美测量获取的值，不包括误差

The score that would have been obtained were measurement perfect and there was no error

- **测量误差** Measurement error

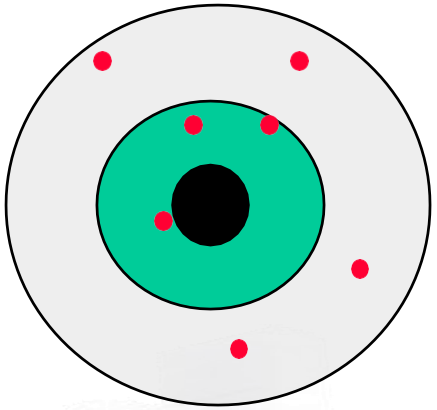
有一些因素而对实际值造成的扭曲程度

Any factor that distorts the true score

Measurement Error

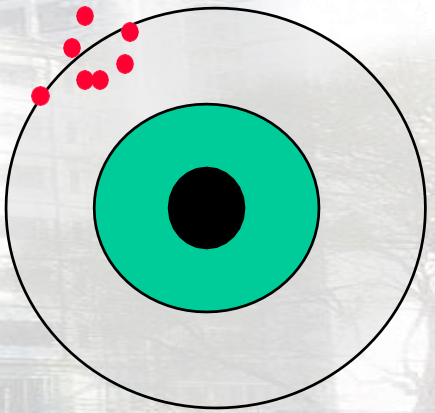
- **随机误差** Random error ("noise")
  - 可能波动-可通过多次测度波动
    Chance fluctuations (tend to cancel out over repeated measurements)
  - 如：情绪，疲劳、动机等

Examples: mood,level of fatigue, motivation, etc.



- **系统误差** Systematic error ("bias")
  - 偏向某个特定方向的波动
    Fluctuations that are slanted in a particular direction
  - 如：引起困惑的术语，错误的刻度

Examples: confusing terms, wrong scales etc.



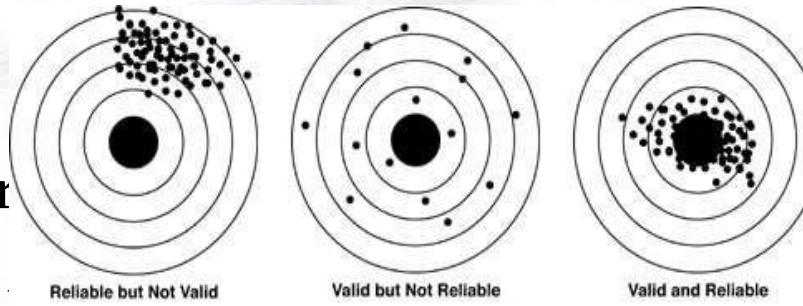
Measurement: Reliability and Validity

- 一个有用的度量，必须即可信又有效
  For a measure to be useful, it must be both reliable and valid
- **可信** Reliable = 度量每次使用时都能得到一致的结果

  Consistent in producing the same results every time the measure is used
- **有效** Valid = 测度了它要测度的事项

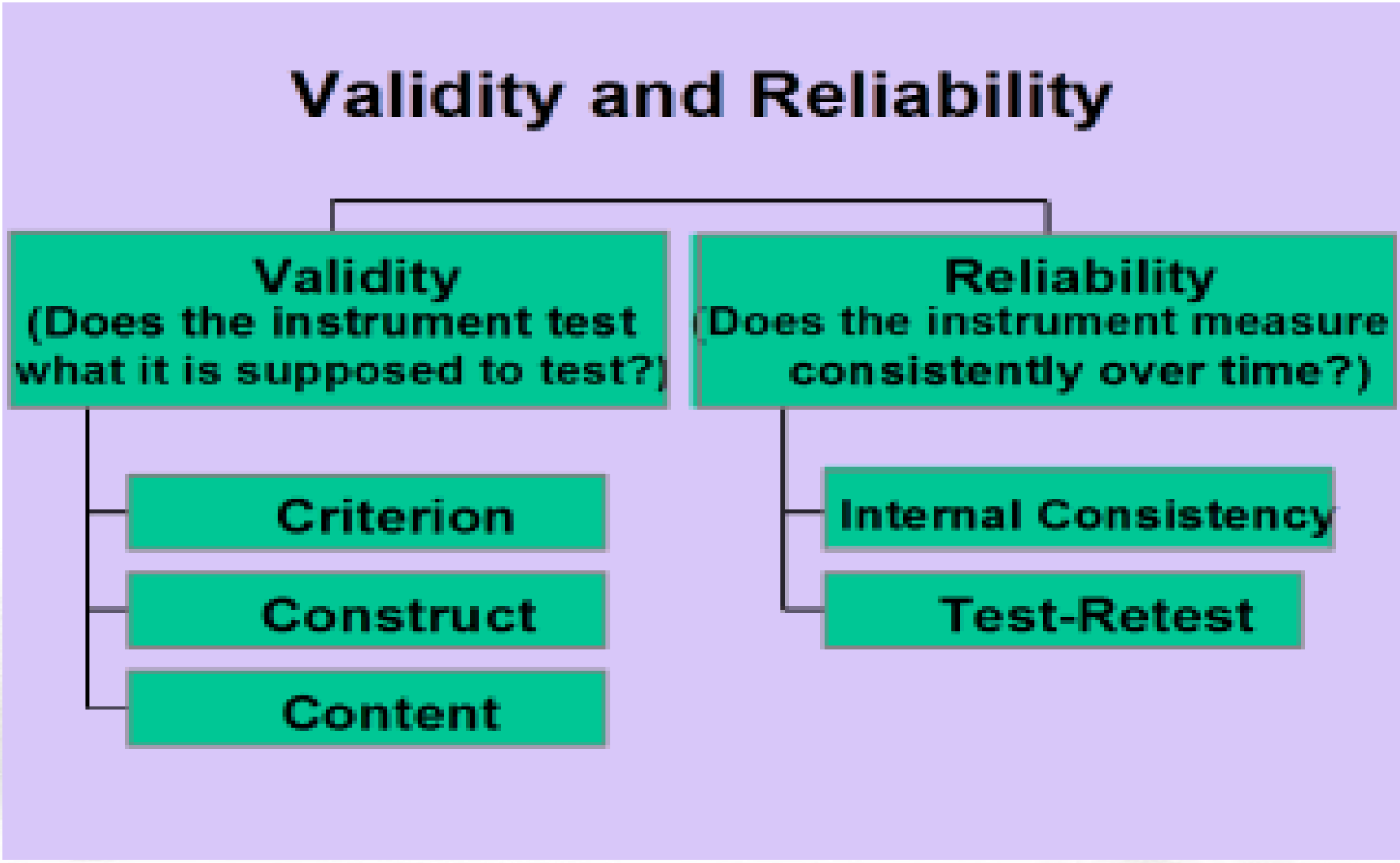
  Measuring what it is supposed to measure

Score = effect + systematic + random error



- 随机误差 Random error: threat to reliability
- 系统误差 Systematic error: threat to validity

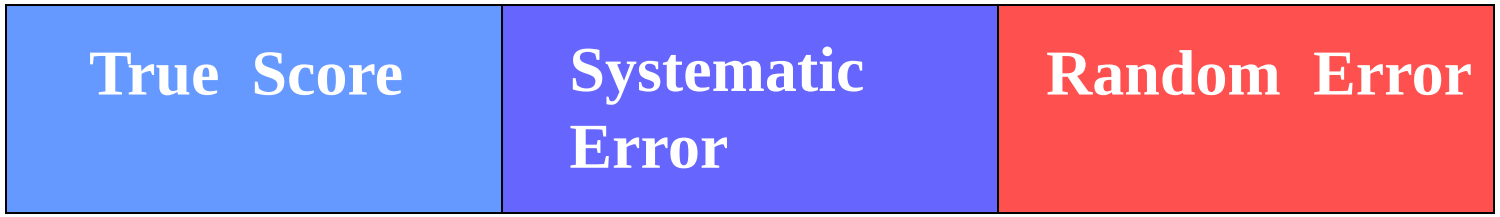
Measurement: Reliability and Validity



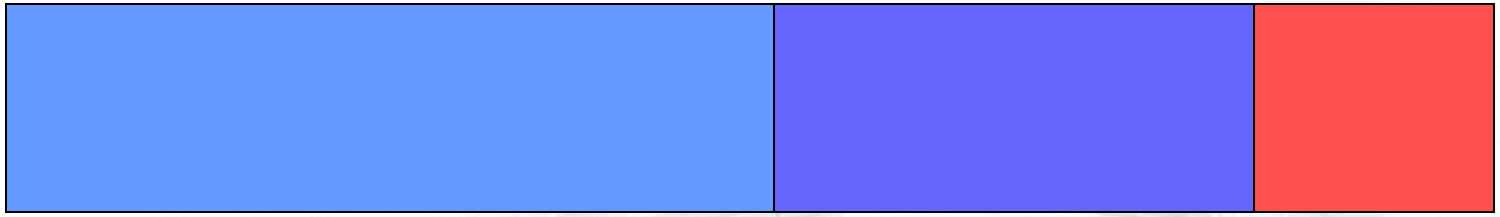
Measurement: Reliability and Validity

- **信度** Reliability （一致、准确、可靠）
  - 度量工具在多大程度上避免了随机误差
    The extent to which the measurement process is free from random errors.
  - 保证了研究结果的一致、准确和可预测性
    Concerned with the consistency, accuracy, and predictability of the research findings

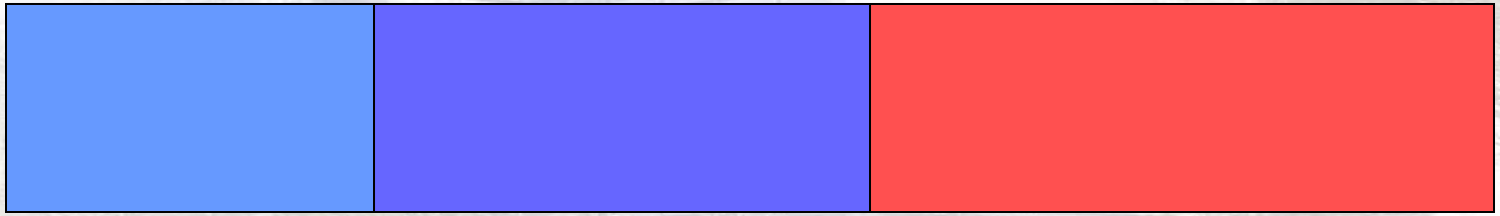
Measurement: Reliability and Validity



More Reliable:



Less Reliable:



Measurement: Reliability and Validity

- **信度类型** Types of reliability
  - 再测信度 Test-retest reliability
  - 副本信度 Equivalent-form reliability
  - 内部一致性 Internal Consistency
    - Split-half 折半信度
    - Cronbach's alpha: average of all possible split-half reliabilities

Types of Reliability

- 再测信度 Test-retest reliability （不同时间，相同测验）

• 在尽可能相同的情况下，使用相同手段再次测度而得到一致结果的能力 the ability of the same instrument to produce consistent results when used a second time under conditions as nearly the same as possible.

- 可通过相关系数进行检验 Quantified with a correlation coefficient
- r值大于0. 7 values are considered good if r >= 0.70
- 再测特定问题或整个调研工具 Test-retest specific questions or the entire survey instrument
- 确保再测在短时间进行 Make sure that you test-retest over very short periods of time

Types of Reliability

- 副本信度 Equivalent-form reliability or alternative-form reliability（不同时间，不同测验）
  - 用尽可能相似的两种手段测量同一物体，得到相似结果的能力 the ability to produce similar results using two instruments as similar as possible to measure the same object.
  - 通过不同用语来测度同种属性 Use differently worded forms to measure the same attribute
  - 问题或应答项进行重述，或顺序变更来开发两个本质相同，表述不同的指标项
  - 重述或顺序调整不能改变指标项的内涵 Change the wording of the response alternatives without changing the meaning

Example: Assessment of depression

Circle one item

Version A:

During the past 4 weeks, I have felt downhearted:

Every day	1
Some days	2
Never	3

Version B:

During the past 4 weeks, I have felt downhearted:

Never	1
Some days	2
Every day	3

Example: Assessment of urinary function

Version A:

During the past week, how often did you usually update your Weibo?

- 1 to 2 times per day
- 3 to 4 times per day
- 5 to 8 times per day
- 12 times per day
- More than 12 times per day

Version B:

During the past week, how often did you usually update your Weibo?

- Every 12 to 24 hours
- Every 6 to 8 hours
- Every 3 to 5 hours
- Every 2 hours
- More than every 2 hours

Types of Reliability

- 内部一致信度 Internal-consistency reliability

• 测量同一属性的不同题项的信度 is a measure of reliability of different survey items intended to measure the same characteristic.

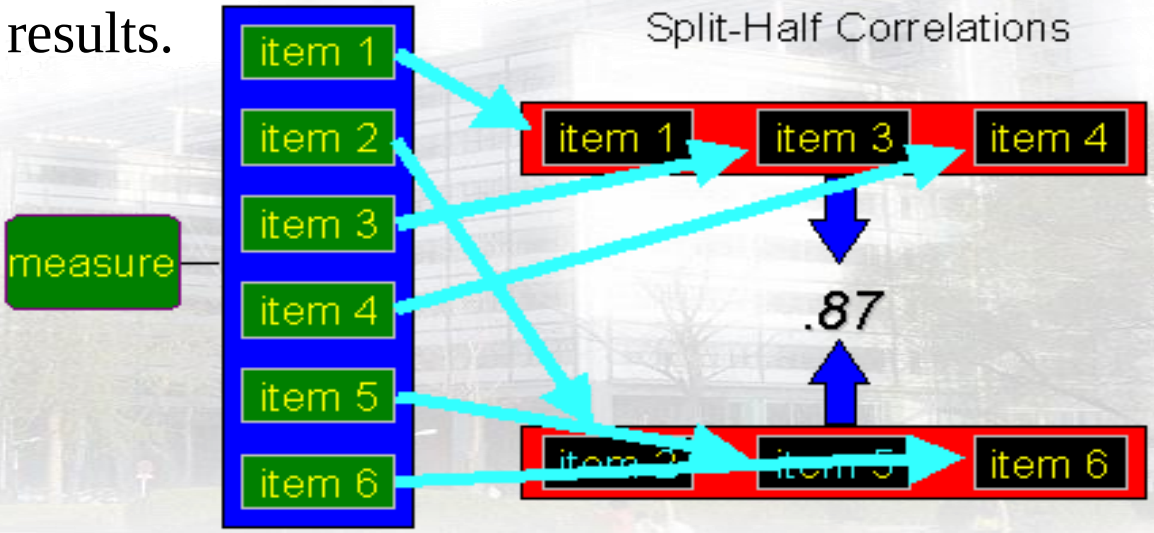
• 不对应单一题项，而对应一组题项 Applied not to one item, but to groups of items that are thought to measure different aspects of the same concept

- Split-half reliability 折半信度 （相同时间，不同测验）相关系数
- Cronbach Alpha 同质性信度

Types of Reliability

- 内部一致信度 Internal-consistency reliability
  - Split-half reliability 折半信度

    （相同时间，不同测验）相关系数
  - 将测度项分成两部分并关联结果的量表信度 the reliability of a scale by dividing into two the total set of measurement items and correlating the results.



Types of Reliability

- 内部一致信度 Internal-consistency reliability
  - 同质性信度 Cronbach Alpha

–测度整合起来测度一个单一构念的一组题项的内部一致性

Measures internal consistency reliability among a group of items combined to form a single scale

–主要反映不同题项如何互补来反映同一构念或属性的不同方面

It is a reflection of how well the different items complement each other in their measurement of different aspects of the same variable or quality

–SPSS, >0.7

Measurement: Reliability and Validity

影响信度的因素 Factors that increase reliability

- 题项数量 Number of items
- 个体间的差异程度High variation among individuals being tested
- 导语的清晰度 Clear instructions
- 理想的测量环境 Optimal testing situation

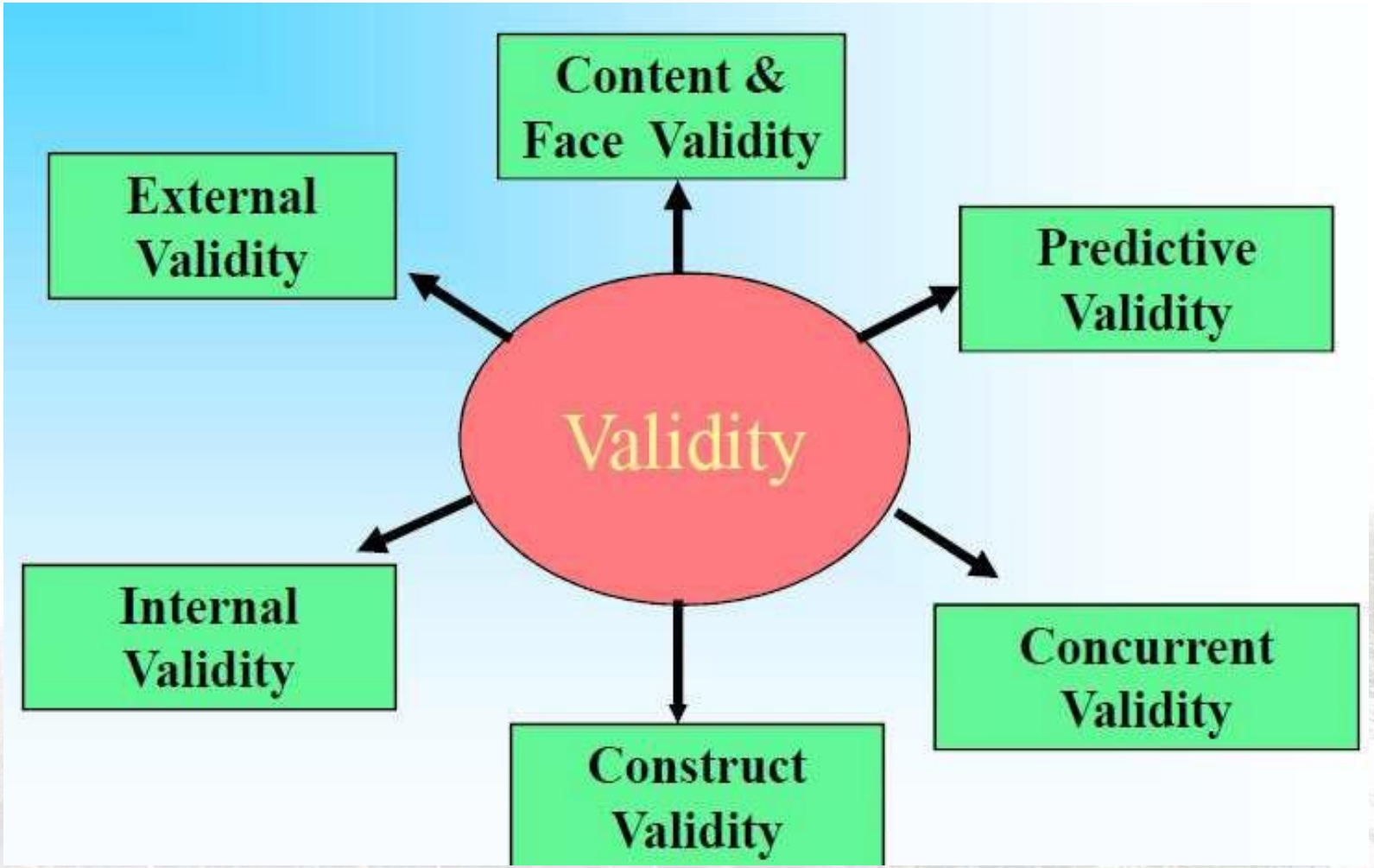
Measurement: Reliability and Validity

- 效度 Validity （真实、准确）
  - 测量过程不受系统和随机误差影响的程度

    the extent to which the measurement process is free from both systematic and random error.
  - 是否测量了我们想要测量的东西

    Are we measuring what we think we are measuring

Research Methodology--Validity



Research Methodology--Validity

- 表面效度 Face validity
- 测度项看起来准确测量目标事物的程度

  The degree to which a measurement seems to measure what it is supposed to measure.

- 方法：专家小组，研究者评判某种测度方法以准确测量目标事物的程度

  Method: expert panel; Researchers judge the degree to which a measurement instrument seems to measure what is supposed to.

Research Methodology--Validity

■ 内容效度 Content validity

• 测量工具体现被研究概念的程度 the degree to which a measurement instrument represent the universe of the concept under study.

• 内容效度是指测度方法内容的代表性或取样准确性 Content validity is the representativeness or sampling adequacy of the content of the measurement instrument.

• 方法：文献、专家组、判别事项 Method; literature, expert panel

■ 内容/表面效度 Content Vs. Face Validity

• 内容效度强调开发量表的过程 Content validity focuses on the procedure we used to develop our measure.

• 表面效度与开发量表之后的事情有关 Face validity is an issue after the measure has been developed.

Research Methodology--Validity

- 效标效度 Criterion-related validity CRV

- 测量工具可以预测指定的效标变量的程度 the degree to which a measurement instrument can predict a variable that is designated a criterion.
- 预测效度 Predictive validity :效标变量的未来水平能被当前量表的测度项（预测变量）预测

Adoption intention vs. Actual adoption

- 同时/并行效度 Concurrent validity : 在与兴趣变量同一时间测度的效标变量可以被测度项（预测变量）预测的程度。

Job performance: self-reported vs. supervisor performance rating

Research Methodology--Validity

- 建构效度/构念效度 Construct validity

• 测量工具能够通过基本理论，在逻辑上连接观测到的现象与构念的程度

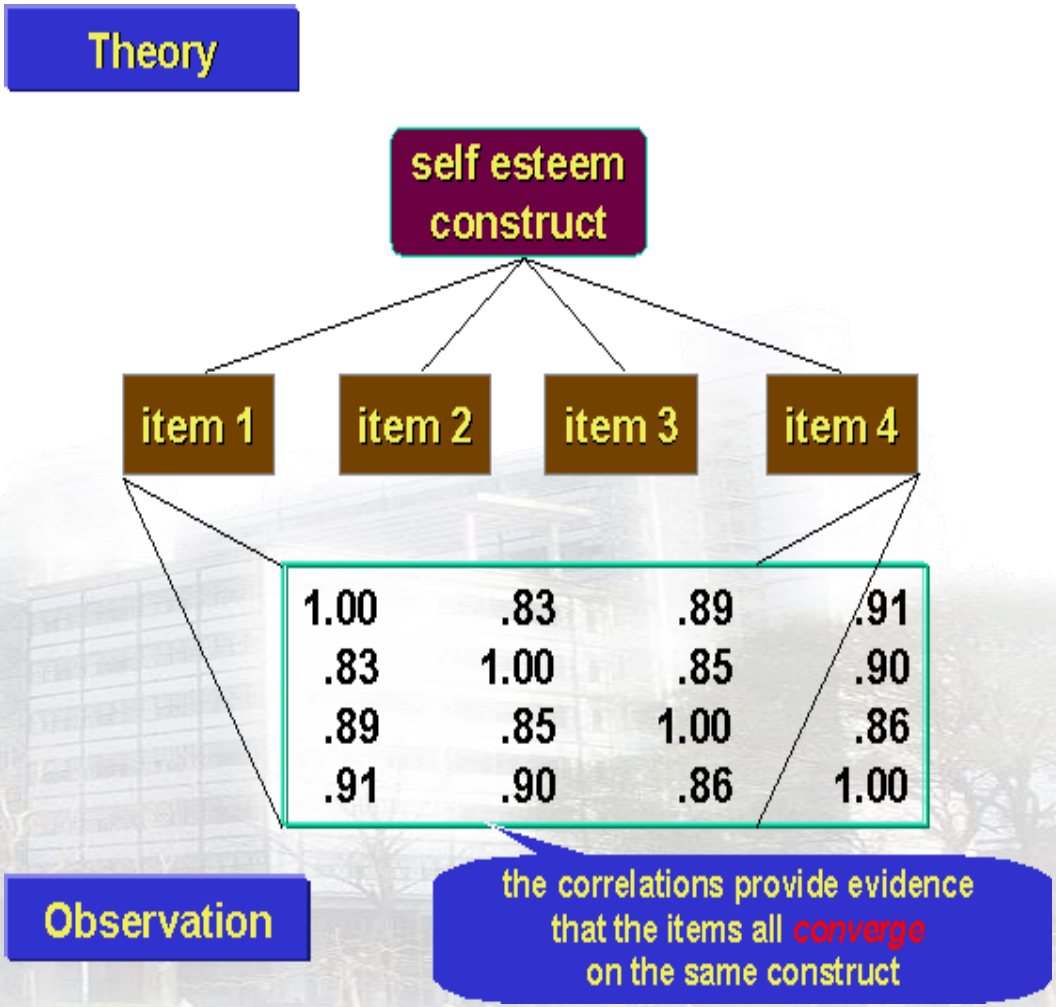
the degree to which a measurement instrument represents and logically connects, via the underlying theory, the observed phenomenon to the construct.

- Convergent validity 收敛效度/聚合效度
- Discriminant validity 区别效度/判别效度

Research Methodology--Validity

- 收敛效度/聚合效度
Convergent validity

  - 旨在测量同一概念的不同测度项间的联系程度
the degree of association among different measurement instruments that purport to measure the same concept.
  - Methods: SPSS, SEM: composite reliability, loading, AVE



Research Methodology--Validity

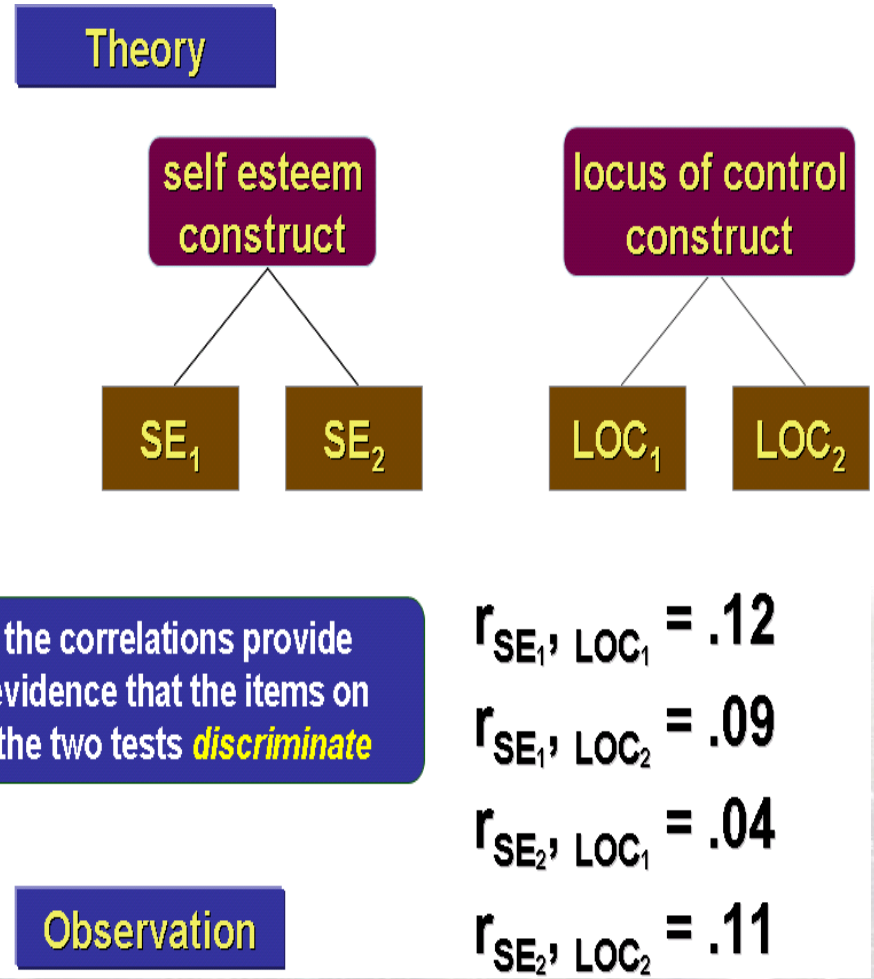
- 区别效度/判别效度

Discriminant validity

- 不同构念之间的低关联程度

the lack of, or low correlation among, constructs that are supposed to be different.

- Methods: SPSS, SEM: cross-loading, chi-square difference, correlation vs. AVE



Convergent vs. Discriminant validity

Theory

self esteem construct

locus of control construct

SE_1 SE_2 SE_3 LOC_1 LOC_2 LOC_3

	SE_1	SE_2	SE_3	LOC_1	LOC_2	LOC_3
SE_1	1.00	.83	.89	.02	.12	.09
SE_2	.83	1.00	.85	.05	.11	.03
SE_3	.89	.85	1.00	.04	.00	.06
LOC_1	.02	.05	.04	1.00	.84	.93
LOC_2	.12	.11	.00	.84	1.00	.91
LOC_3	.09	.03	.06	.93	.91	1.00

Observation

the correlations support both *convergence* and *discrimination*, and therefore *construct* validity

Research Methodology--Validity

Type	What Is Measured	Methods
Content	Degree to which the content of the items adequately represents the universe of all relevant items under study.	• Judgmental • Panel evaluation with content validity ratio
Criterion-Related	Degree to which the predictor is adequate in capturing the relevant aspects of he criterion.	• Correlation
Concurrent	Description of the present; criterion data are available at the same time as predictor scores.	• Correlation
Predictive	Prediction of the future; criterion data are measured after the passage of time.	• Correlation
Construct	Answers the question, "What accounts for the variance in the measure?"; attempts to identify the underlying construct(s) being measured and determine how well the test represents it (them).	• Judgmental • Correlation of proposed test with established one • Convergent-discriminant techniques • Factor analysis • Multitrait-multimethod analysis

Research Methodology--Validity

- 内部效度 Internal Validity
  - 自变量变化对于因变量的影响程度。如：除了自变量本身，存在其他因素影响自变量与因变量的关系吗？

    The extent to which your manipulation of the IV actually influences the "scores" on the DV., i.e., could something else have affected your IV-DV relationship other than the IV itself?
  - 在混杂变量得以适当控制的情况下，实验研究通常具有高内部效度

    Experimental studies generally have high internal validity if confounding variables are properly controlled.

Research Methodology--Validity

- 外部效度 External Validity
  - 研究的结果适用于研究以外的其他人、组织或事件的程度
    The extent to which the results of the study are generalizable to other people, organization, or events outside the study.
  - 外部效度不仅仅受测度影响（其他因素如样本量、样本类型、实验设计）
    It is influenced by more than just measurement (e.g., sample size, type of sample, design of experiment),
  - 有效的测度是外部效度的必要非充分条件
    Valid measurement is a necessary but not sufficient condition for external validity.

Table 3
Loadings of the measures

Constructs	Items	Loading	Standard error	*t*-statistic	Composite reliability	AVE
Coercion Power	CoPo1	0.872	0.024	36.796		
	CoPo2	0.903	0.020	45.554	0.893	0.736
	CoPo4	0.796	0.037	21.269		
Reward Power	*RePo1*	*0.642*	*0.078*	*8.199*		
	RePo2	0.848	0.041	20.765	0.834	0.630
	RePo3	0.873	0.023	38.427		
Legitimate Power	*LePo1*	*0.602*	*0.102*	*5.927*		
	LePo2	0.764	0.053	14.462	0.766	0.525
	LePo3	0.793	0.039	20.483		
Information Power	InPo1	0.725	0.073	9.969		
	InPo2	0.815	0.039	21.001	0.820	0.603
	InPo3	0.787	0.047	16.712		
Expert Power	ExPo1	0.846	0.026	33.123		
	ExPo2	0.813	0.063	12.944	0.853	0.660
	ExPo3	0.778	0.036	20.787		
Referent Power	RfPo1	0.881	0.025	35.559	0.874	0.777
	RfPo2	0.881	0.025	35.559		
Cognitive Trust	Cotr1	0.747	0.053	14.047		
	Cotr2	0.850	0.027	31.179	0.855	0.664
	Cotr3	0.844	0.032	26.558		
Affective Trust	Aftr1	0.861	0.024	36.685		
	Aftr2	0.878	0.031	28.459	0.886	0.721
	Aftr3	0.807	0.043	18.737		
Customer Pressures	CuPr1	0.834	0.033	24.885		
	CuPr2	0.810	0.037	21.590	0.868	0.687
	CuPr3	0.843	0.037	22.680		
Supplier Pressures	SuPr1	0.760	0.058	13.040		
	SuPr2	0.859	0.036	23.849	0.865	0.682
	SuPr3	0.855	0.031	26.983		
Normative Pressures	NoPr1	0.825	0.029	28.563		
	NoPr2	0.823	0.028	29.116	0.840	0.636
	NoPr3	0.743	0.037	20.044		
eSCMS	AdIn1	0.891	0.029	31.105		
Adoption Intention	AdIn2	0.921	0.021	44.285	0.928	0.811
	AdIn3	0.890	0.023	39.539		
Mimetic Pressures	Mipr1	0.777	0.036	21.553		
	MiPr2	0.878	0.025	34.763	0.871	0.692
	MiPr3	0.839	0.039	21.542		
Dependence	Dep1	0.879	0.024	36.571		
	Dep2	0.863	0.040	21.681	0.900	0.750
	Dep3	0.856	0.026	33.191		

Properties of Measures

Table 4
Assessment of discriminant validity

Constructs	CoPo	RePo	LePo	InPo	ExPo	RfPo	CoTr	AfTr	NoPr	CuPr	SuPr	AdIn	Dep	MiPr	Size
Coercion Power (CoPo)	0.86														
Reward Power (RePo)	0.39	0.79													
Legitimate Power (LePo)	0.49	0.37	0.73												
Information Power (InPo)	0.29	0.32	0.36	0.78											
Expert Power (ExPo)	0.35	0.28	0.45	0.53	0.81										
Referent Power (RfPo)	0.10	0.15	0.34	0.51	0.44	0.88									
Cognitive Trust (CoTr)	−0.12	0.02	−0.04	0.26	0.28	0.26	0.82								
Affective Trust (AfTr)	−0.09	0.06	0.09	0.41	0.26	0.37	0.63	0.85							
Normative Pressures (NoPr)	0.14	0.09	0.06	0.23	0.26	0.06	0.09	0.08	0.79						
Customer Pressures (CuPr)	0.14	0.17	0.11	0.27	0.25	0.15	0.13	0.15	0.68	0.83					
Supplier Pressures (SuPr)	0.09	0.18	0.12	0.32	0.34	0.25	0.11	0.09	0.62	0.80	0.83				
Adoption Intention (AdIn)	−0.08	−0.09	−0.07	0.09	0.11	0.13	0.26	0.22	0.50	0.45	0.53	0.90			
Dependence (Dep)	−0.02	0.00	0.08	0.17	0.13	0.14	0.13	0.08	−0.01	−0.09	0.06	−0.06	0.83		
Mimetic Pressures (MiPr)	0.11	0.04	0.08	0.20	0.24	0.15	0.05	0.05	0.74	0.72	0.75	0.50	−0.04	0.87	
Firm Size (Size)	0.08	0.16	0.18	0.09	0.20	0.07	0.13	0.25	0.06	−0.02	−0.01	0.04	0.04	0.05	NA

Note: The diagonal elements are the square roots of AVEs.


Q&A


THANK YOU