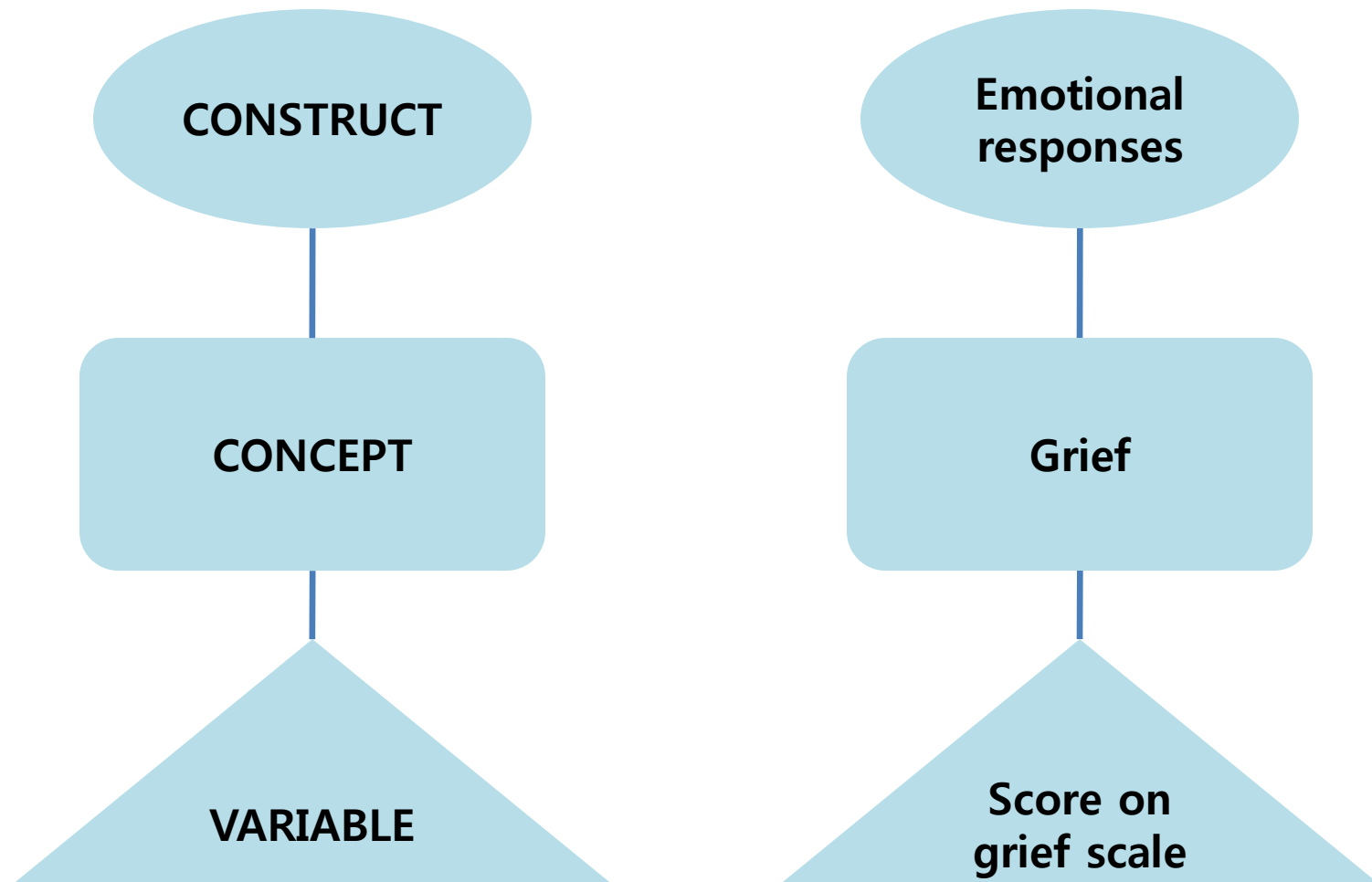# 측정도구 선택

이주희, PhD, GNP, RN
2014. 8. 25

# Contents

- Brief explanation on instrumentation & measurement
- Reliability
- Validity
- Another considerations

# Construct-Concept-Variable

CONSTRUCT

CONCEPT

VARIABLE

Emotional responses

Grief

Score on grief scale

# Instrumentation

- The process whereby one decides how to measure a variable

- The process of selecting or developing tools and methods appropriate for measuring an attribute or characteristic of interest.

# Measurement

- The process of assigning numbers to objects to represent the kind and/or amount of attributes or characteristics possessed by those objects.

# Concept of measurement error

**Observed score** = **True score** + **Error**

# Concept of measurement error

- *Systematic error*

: affects all observations equally and be a **constant error** or affect certain types of observations differently than others and be a **bias**
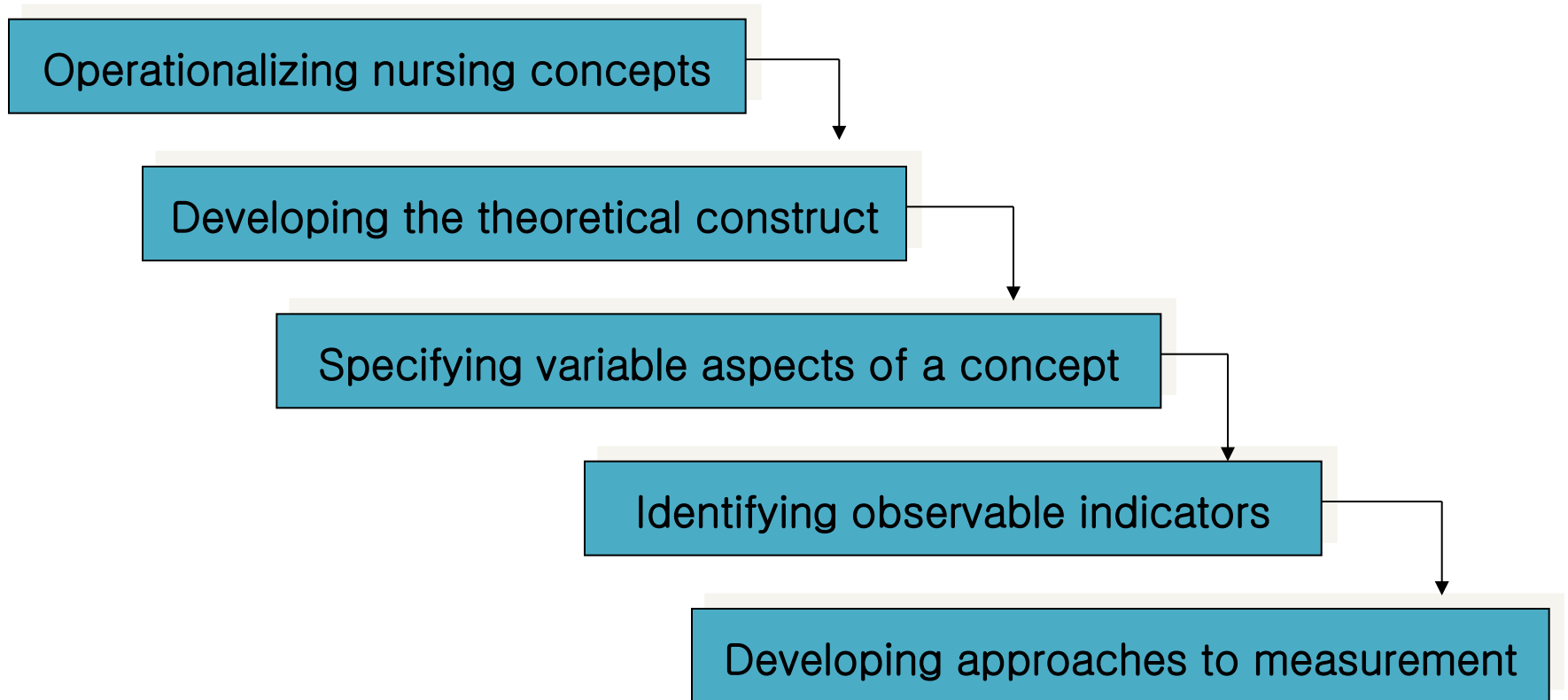
- *Random error*

: limit the degree of lawfulness in nature by complicating relationships

: also influence measurements in the behavioral sciences

# Measurement in science

1. Represent quantities of attributes numerically

2. Define whether the objects fall in the same or different categories with respect to a given attribute

- Measurement always concerns some particular feature of objects. One cannot measure objects, one measures their attributes.

- Therefore, measurement requires a process of abstraction & measurement makes us consider the nature of an attribute carefully before attempting measurement.

Operationalizing nursing concepts

Developing the theoretical construct

Specifying variable aspects of a concept

Identifying observable indicators

Developing approaches to measurement

# Advantages of Standardized Measures

- Objectivity
- Quantification
- Communication
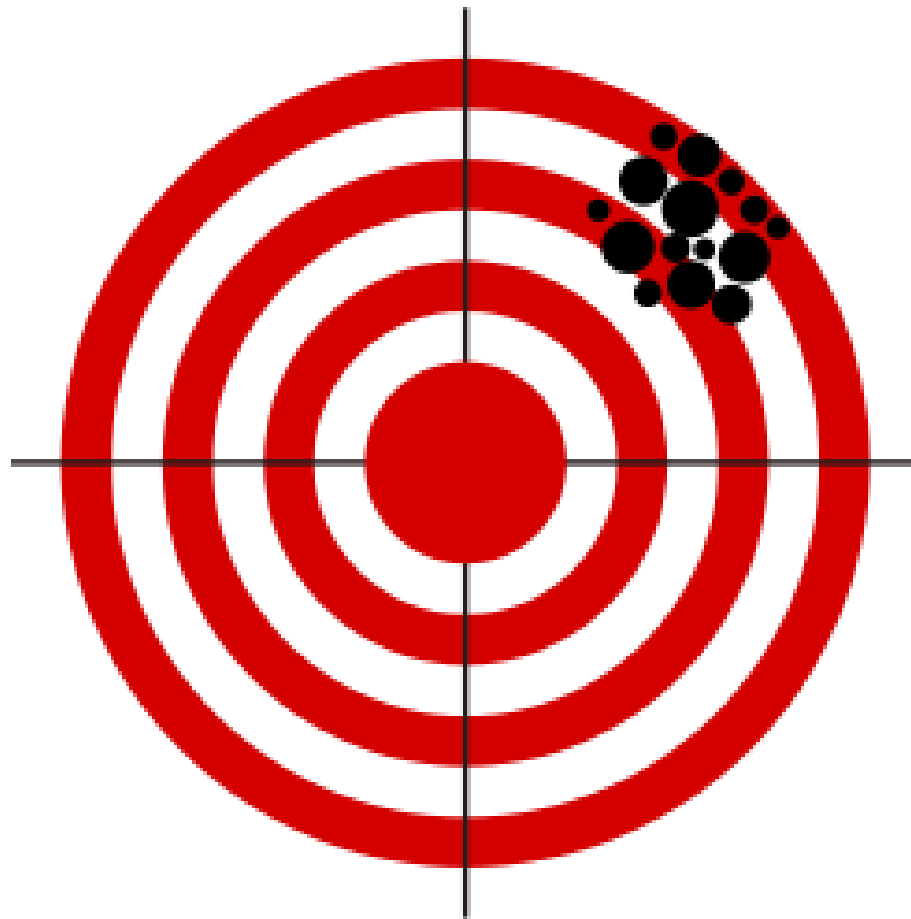- Economy
- Scientific generalization
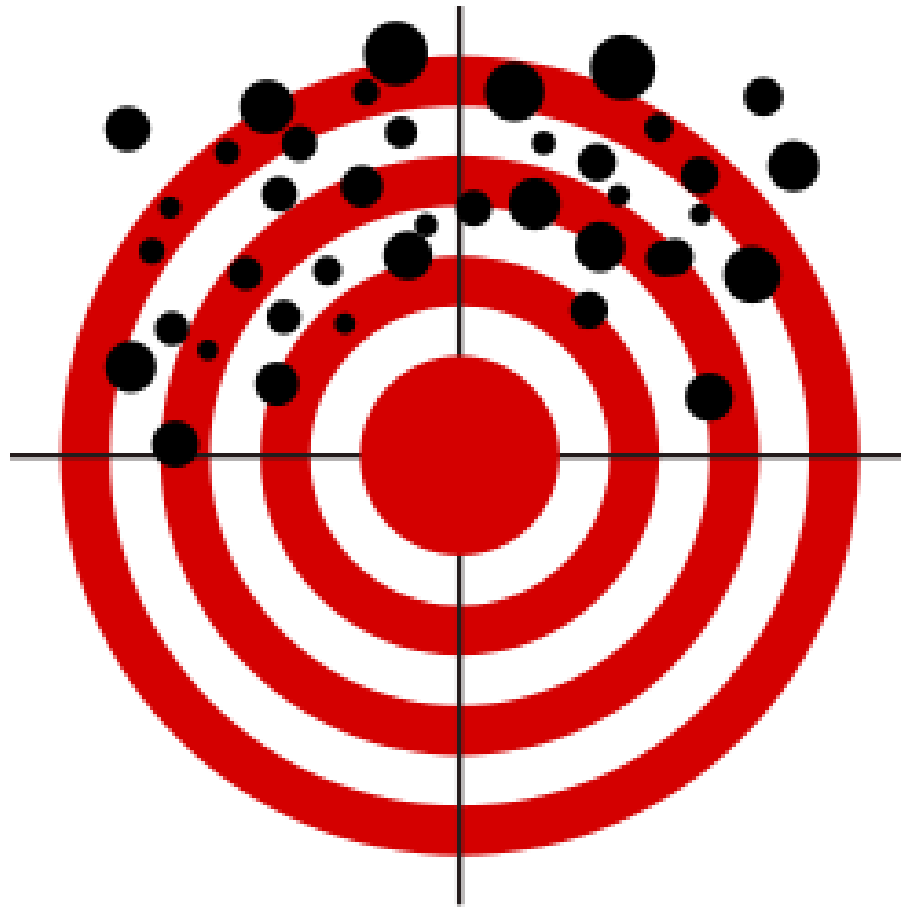
# Reliability & Validity

- **Reliability** refers to the *reproducibility of assessment data or scores, over time or occasions.*


- **Validity** refers to the *degree of meaningfulness for any interpretation of a test score.*
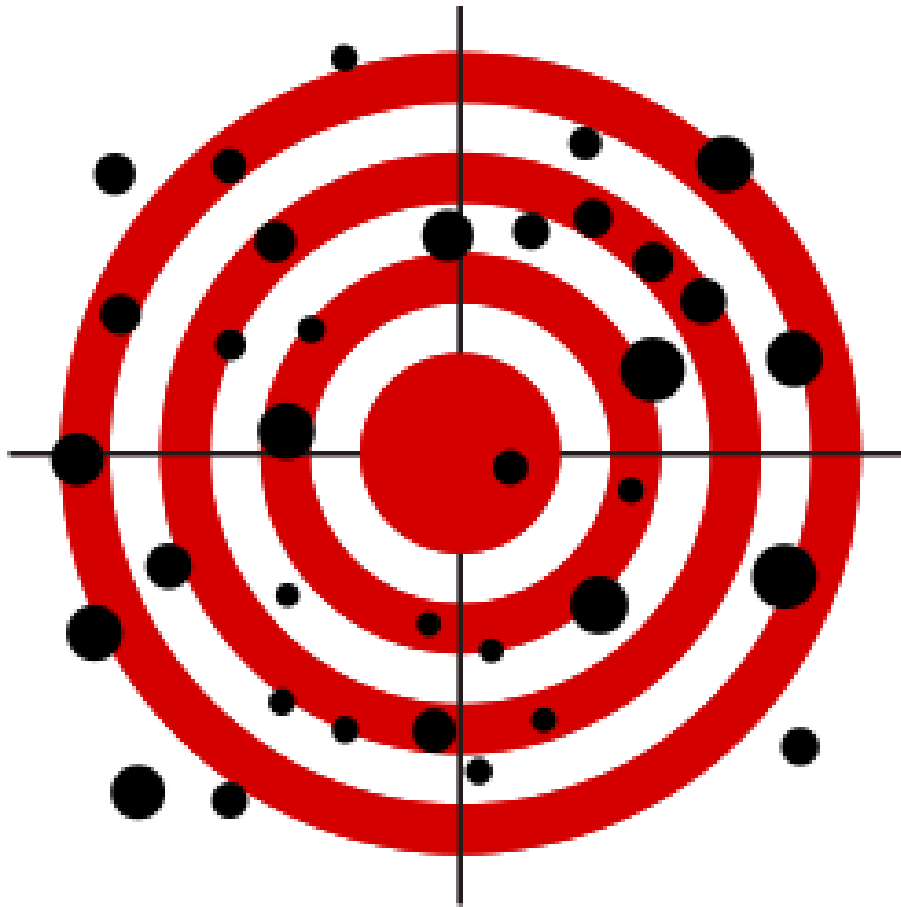
**Both Reliable & Valid**

Reliable, Not Valid

Unreliable & Unvalid

Unreliable, But Valid

# Reliability

- Reliability is a correlation computed between two events.
  1. Stability: extent to which an instrument yield the same results on repeated administrations
  2. Internal consistency: extent to which all the items are measuring the same attribute
  3. Equivalence: estimates of inter-rater reliability
- Reliability assessments involve computing a reliability coefficient.

# Statistical definition of reliability

$$= \frac{Var(T)}{Var(O)}$$

*The reliability coefficient
can be shown to estimate the ratio of
variance in true scores
to the variance in observed scores*

# Stability

- Stability: extent to which an instrument yield the same results on repeated administrations

- Test-retest reliability

- Estimation based on the correlation between two (or more) administrations of the same item, scale, or instrument for different occasions.

- Appropriate for relatively enduring attributes

# Internal consistency

- Internal consistency: extent to which all the items are measuring the same attribute
- Administer to one group on one occasion
- Split-half technique
- Coefficient alpha(Cronbach's alpha)
  - The degree of items' intercorrelations
  - The proportion of variability that is shared among items

# Coefficient Alpha

- Indicator of performance on any other item in the same instrument

- alpha= $(k/(k-1)) * [1- (s2i)/s2sum]$

k = number of items on the measure
s2i = the sum of the individual item variances
sS2sum = the variance of the distribution of test scores

# Alpha coefficient

- *Cut-off criteria*
  - .60 is common in exploratory research
  - adequate scale; at least ≥ .70
  - good scale;  ≥ .80
  - In SPSS, Cronbach's alpha is found under Analyze> Scale> Reliability Analysis

# KR20 / KR21

- Internal  consistency with dichotomous data

KR20= (K/(K-1)) * [1-  (∑pq)/s2test]

     K = number of items on the measure
     p = the proportion of correct responses to an item
     q = the proportion of incorrect or missing responses
    s2test = the variance of the test-score distribution
    pq=the variance for each item
    ∑pq=the sum of the item variances

# Equivalence

- Equivalence: estimates of inter-rater reliability
- Cohen's Kappa

$$\frac{\text{The number of agreement}}{\text{The number of agreement} + \text{The number of disagreement}}$$

  - .60 : minimally acceptable
  - >.75 : good

# Validity

1. Construct validity: psychological attributes, trait

2. Content validity: relevance, representativeness

3. Criterion-related validity (predictive and concurrent): establishing a statistical relationship with a particular criterion /empirical validity

☞ Face validity

# Construct validity

- **Construct**: a variable is abstract and latent, which does not exist as an observable dimension of behavior

*Scientists cannot do without constructs!*

SCIENTIFIC GENERALIZATION

# 3 aspects of construct validation

- Specifying well the domain of observables related to the construct (**outlining the domain**)

- Determining the extent to which observables tend to measure the same thing, several different things or many different things from empirical research and statistical analyses (**investigating relations among different proposed measures of a construct**)

- Performing subsequent individual differences studies and/or experiments to determine the extent to which supposed measures of the construct are consistent with best guesses about the construct!

- To determine construct validity, a measure must fit a theory about the construct: but to use this evidence, one must assume **the theory is true**.
- Constructs A and B correlate positively.
- X is a measure of construct A.
- Y is a measure of construct B.
- X and Y correlated positively.

# Construct validity

*Test construction: affect measure*

- Known groups or contrasted groups approach

- Hypothesis testing approach

- Multi-Trait Multi-Method approach

- Factor analysis

# Content validity

- The degree to which an instrument has an appropriate items for the construct being measured.

- Content validity index(CVI): evaluated by expert

# CVI

- 4-point scale of relevance
  - From 1(not relevant) to 4(highly relevant)
- Item CVI

$$\frac{\text{The number of experts giving a rating 3 or 4}}{\text{The number of experts}}$$

  - 0.8 : acceptable

# Criterion-related validity

- The degree to which the instrument correlates with an external criterion.

- Validity coefficient is correlation coefficients between the new scale and the gold standard or future outcome.

    1. Concurrent validity
    2. Predictive validity

# Another consideration…

## *The number of items in an instrument*

- Shorter scales are good?
- Longer scales are good?

-Scale length affects the precision of alpha

- Shorter scales

: Place less of a burden on respondents,

- Longer scales

: Tend to be more reliable. It has been noted that alpha increases when more items are included (unless they are relatively poor items).

# Tips

- Popular tools
- Population specific tools
- Tools with cut-off point

# Example

- MMSE-K

; ≤20=generally suggests dementia but may also be found in acute confusion, schizophrenia or severe depression


; ≤24=may indicate dementia in some patients who are well educated and who do not have any of the above conditions

# Thank you