MEASUREMENT ERROR

PSY443
TESTING AND ASSESSMENT IN PSYCHOLOGY

WHAT IS MEASUREMENT ERROR?

- Measurement errors -situations either related to the research or not!!

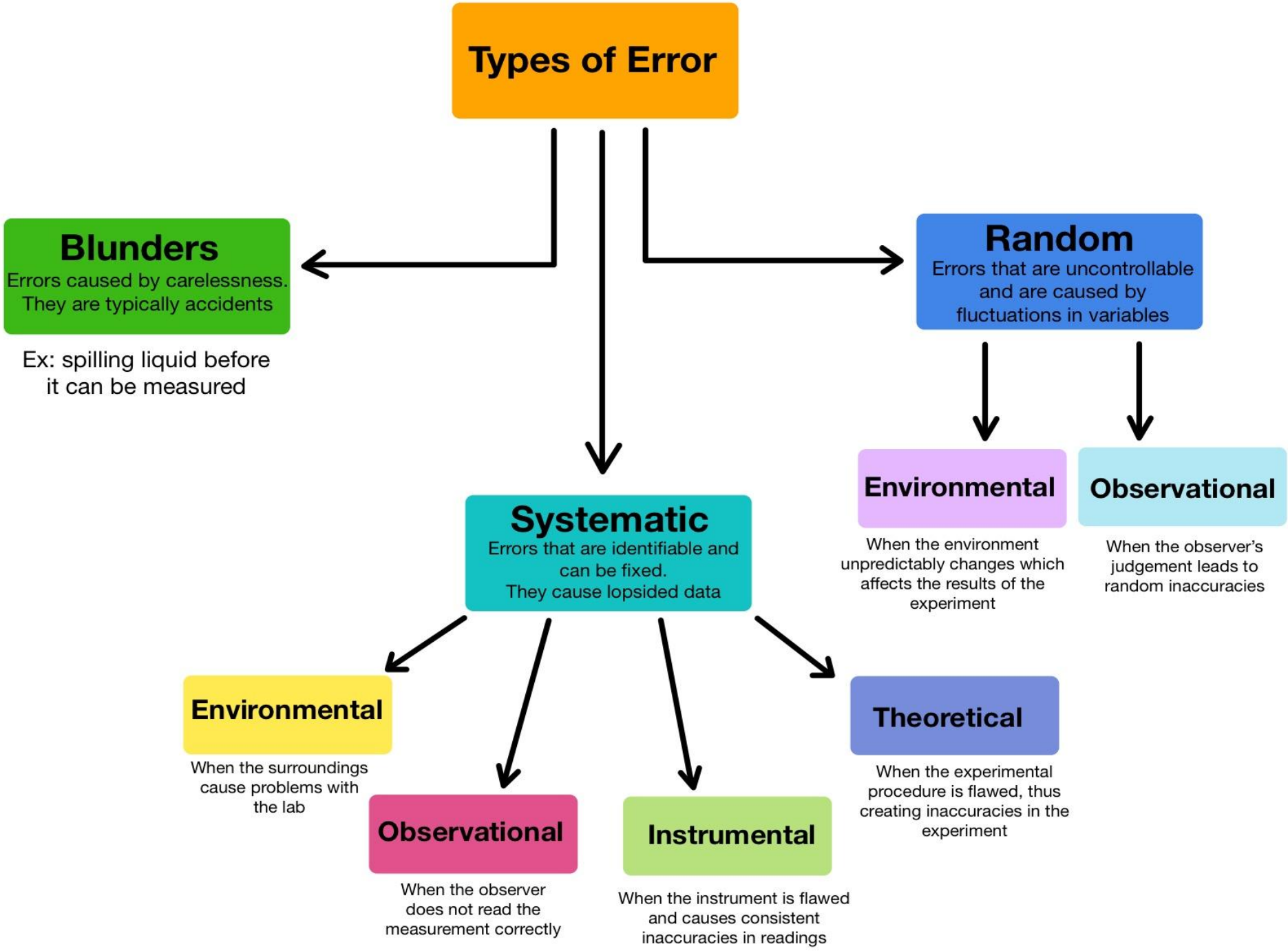
Research-related errors include issues;

- Poorly designed questions
- Unclear expression
- Incorrect sample selection,
- Miscoding of data (Lin, 1976).

WHAT ELSE? - MEASUREMENT ERROR

Errors which are not related to research;

- **Social historical events**
- **Personal historical events**
- **Natural maturation processes** (Lin, 1976)


Types of Error
Blunders
Errors caused by carelessness. They are typically accidents
Ex: spilling liquid before it can be measured
Random
Errors that are uncontrollable and are caused by fluctuations in variables
Environmental
When the environment unpredictably changes which affects the results of the experiment
Observational
When the observer's judgement leads to random inaccuracies
Systematic
Errors that are identifiable and can be fixed.
They cause lopsided data
Environmental
When the surroundings cause problems with the lab
Observational
When the observer does not read the measurement correctly
Instrumental
When the instrument is flawed and causes consistent inaccuracies in readings
Theoretical
When the experimental procedure is flawed, thus creating inaccuracies in the experiment

TYPES OF MEASUREMENT ERROR

TWO MAIN TYPES:

Systematic Errors: These are consistent and repeatable errors - caused by factors such as poorly designed questions, or misinterpretation of responses.

- It affects **validity.**
- It does **not affect reliability.**

Example: A teacher who tells students that they will receive an extra 5 points for behaving well in class…

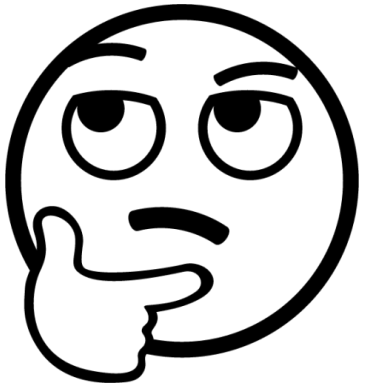
Random Errors: These are unpredictable - respondents' understanding, or timing.

- It directly affects **validity.**
- It indirectly affects **reliability.**

Example: Careless scoring without awareness…

SOURCES OF ERROR

1. The individual conducting the measurement.
2. The measurement tool.
3. The environment in which the measurement is conducted.
4. The individual being measured.



WHAT ABOUT ERRORS caused by the measurement tool??

TRUE SCORE THEORY (CTT)

- CTT **(Klasik Test Kuramı)** - measurement structure used to understand, organize, and interpret measurement results…
- Every measurement include - **a certain amount of error.**
- The measurement results - three components: **True Score (T), Observed Score (X), and Error.**

MEASUREMENT ERROR

- **True Score (T):** The ideal score- the measurement tool - completely reliable.
- **Observed Score (X):** The score that the sum of the true score and the errors

THE FORMULA:

$$X = T + E$$

- **Error (E):** Represents the uncertainties in the measurement. Errors can be systematic or random…

MEASUREMENT ERROR

- The standard error of measurement - the distribution of scores obtained from a measurement tool.
- An important indicator for evaluating the reliability of a test or survey.

Standard Error (SE): how much **the average score** obtained from a sample deviates from the population average.
- helps to understand - the sample mean reflects the population mean.

$\mathbf{S_e}$: Standard error
$\mathbf{S_x}$: Standard deviation
$\mathbf{r_x}$: Reliability

$$S_e = S_x\sqrt{1 - r_x}$$

MEASUREMENT ERROR

The true average of scores is:

- With a 68% probability, within the range of (-, +) **1SE**
- With a 95% probability, within the range of (-, +) **2SE**
- With a 99% probability, within the range of (-, +) **3SE**

MEASUREMENT ERROR

example: In a research method class mid-term exam, the average score of the 2nd year psychology students is 75, the standard deviation of the scores is 2, and the reliability of the test is calculated as 0.75.

- What are the ranges of the true averages for the psychology students in this class with 68%, 95%, and 99% probability?

- Standard deviation ($\mathbf{S_x}$): 2
- Reliability ($\mathbf{r_x}$): 0.75
- **Standard error** ($\mathbf{S_e}$): =?

$$S_e = S_x\sqrt{1 - r_x}$$

MEASUREMENT ERROR

68% Probability Range:

- Lower limit: 75−1=74
- Upper limit: 75+1=76

Range: (74,76)

95% Probability Range:

- Lower limit: 75−2=73
- Upper limit: 75+2=77

Range: (73,77)

99% Probability Range:

- Lower limit: 75−3=72
- Upper limit: 75+3=78

Range: (72,78)

- **68% Probability Range**: (74,76)
- **95% Probability Range**: (73,77)
- **99% Probability Range**: (72,78)

WHAT IS RELIABILITY?

WHAT IS RELIABILITY?

- Reliability - consistency and repeatability of a measurement tool (test, survey, assessment etc.).
- A reliable measurement tool - similar results…
- Reliability - crucial criterion for determining <u>how accurate and dependable</u> your measurement tool/assessment!

COMPONENTS OF RELIABILITY

- **Consistency**: Getting similar results when applied - different times or different samples.
- **Internal Consistency**: The items/questions in the measurement tool correlation - with each other (Cronbach's alpha).
- **Test-Retest Reliability**
- **Parallel Forms Reliability**

TYPES OF RELIABILITY

1. **Test–Retest Reliability**
2. **Parallel Forms Reliability**
3. **Inter-rater or Inter-observer reliability**
4. **Internal consistency reliability**

TYPES OF RELIABILITY

TEST-RETEST RELIABILITY

- Test-retest reliability – obtained scores from the measurement tool or assessment should consistent time
- **It evaluates how stable and reliable the results are when the same test is administered to the same group of people on two different occasions.**
- For instance, Happiness Scale (x2) ☺

TYPES OF RELIABILITY

PARALLEL FORMS OF RELIABILITY

- **Alternate forms reliability** - measure of the consistency between two different versions of the same test or assessment.
- It evaluates whether different forms of a test **produce similar results** when administered to **the same group** of individuals.
- One of the most widely used method in Psyc.

TYPES OF RELIABILITY

INTER-RATER RELIABILITY

- Consistency between **different raters or observers** who evaluate the same situation…
- It is an important aspect of research and assessment- Subjective judgments are involved.

- **Cohen's Kappa (K)**

COHEN'S KAPPA (K)

- A statistical measure that accounts for agreement occurring by chance.
- It is commonly used for categorical data…

The Formula: $$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

RELIABILITY

INTERNAL CONSISTENCY

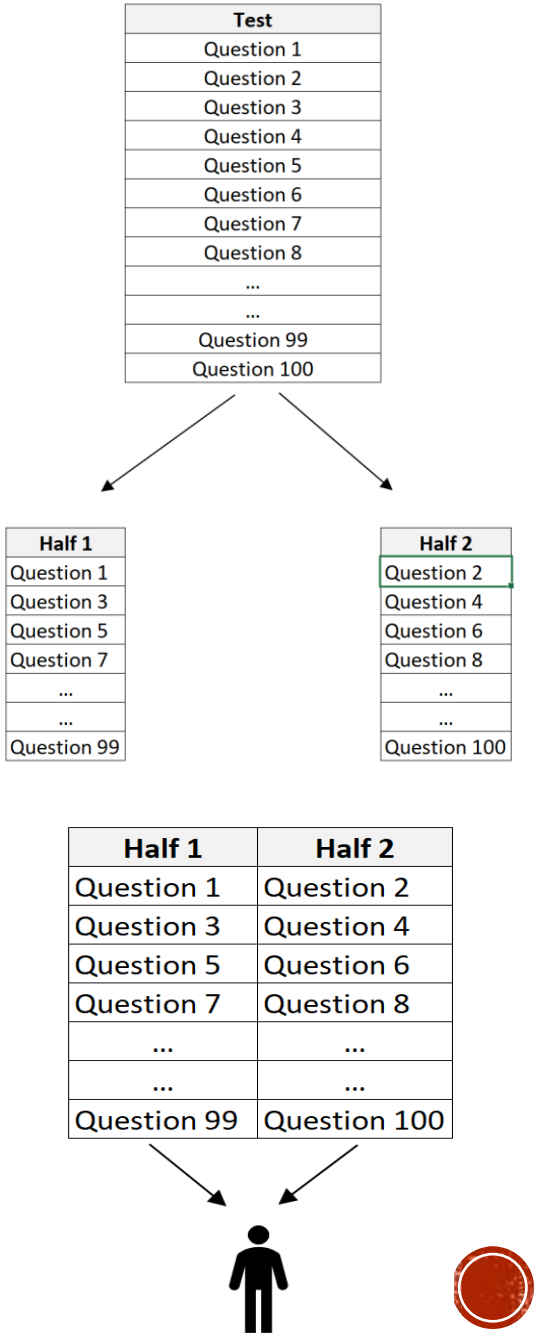
Internal consistency (iç tutarlılık)- The items in a measurement tool, tests or survey are consistent in measuring the same sit. or trait.

Types of internal consistency are;

1. **Split-Half Coefficient**
2. **Kuder-Richardson Approach**
3. **Alpha Coefficient**

SPLIT-HALF COEFFICIENT

- **Split-half coefficient** is a method used to assess the internal consistency of a test or measurement tool.
- It evaluates the reliability of a test by splitting it into two equal halves and comparing the scores from each half.
- This method involves dividing a test into two parts (usually randomly) and then correlating the scores from each part.
- This helps to determine if both halves of the test produce consistent results.



THE KUDER-RICHARDSON

- The Kuder-Richardson - used to assess the reliability of assessments, tests.
- Dichotomous items (eg. right/wrong, yes/no)
- **KR-20**
- Values 0 to 1- higher values greater reliability…
- Reliability value >0.70 is acceptable!!!

CRONBACH'S ALPHA COEFFICIENT

- Cronbach's Alpha - the internal consistency in a questionnaire.
- Multiple items - measure the same underlying construct (attitudes, traits etc.)
- The coefficient ranges from 0 to 1.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.755	.768	9

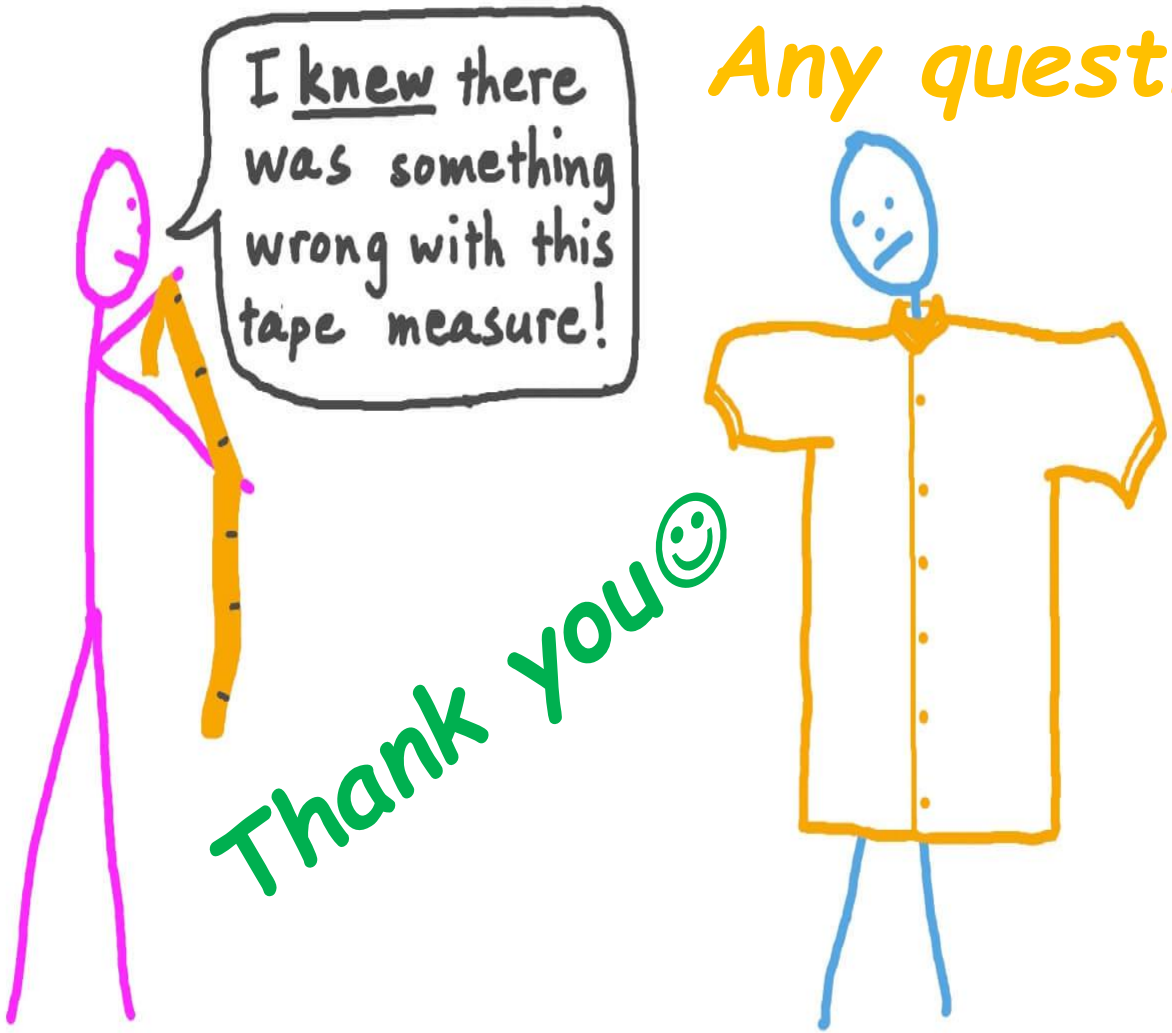
CRONBACH'S ALPHA COEFFICIENT

The Formula

$$\alpha = \frac{N \cdot \bar{c}}{\bar{v} + (N-1) \cdot \bar{c}}$$

Cronbach's Alpha Value	Reliability Coefficient
$\alpha > 0.9$	Excellent
$\alpha > 0.8$	Good
$\alpha > 0.7$	Acceptable
$\alpha > 0.6$	Questionable
$\alpha \geq 0.5$	Poor

MEASUREMENT ERROR



Any questions???

Thank you☺