

# Reliability

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# What is reliability for a test?

Britt, M. (2012). Test Reliability Explained

<http://www.youtube.com/watch?v=kDW0Y5ms>

[BOA](#)

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# Test Reliability Recap

- Definition

**Consistency** of measurements when the testing procedure is **repeated** on a population of **individuals or groups** (Standards for Educational and Psychological Testing, 1999).

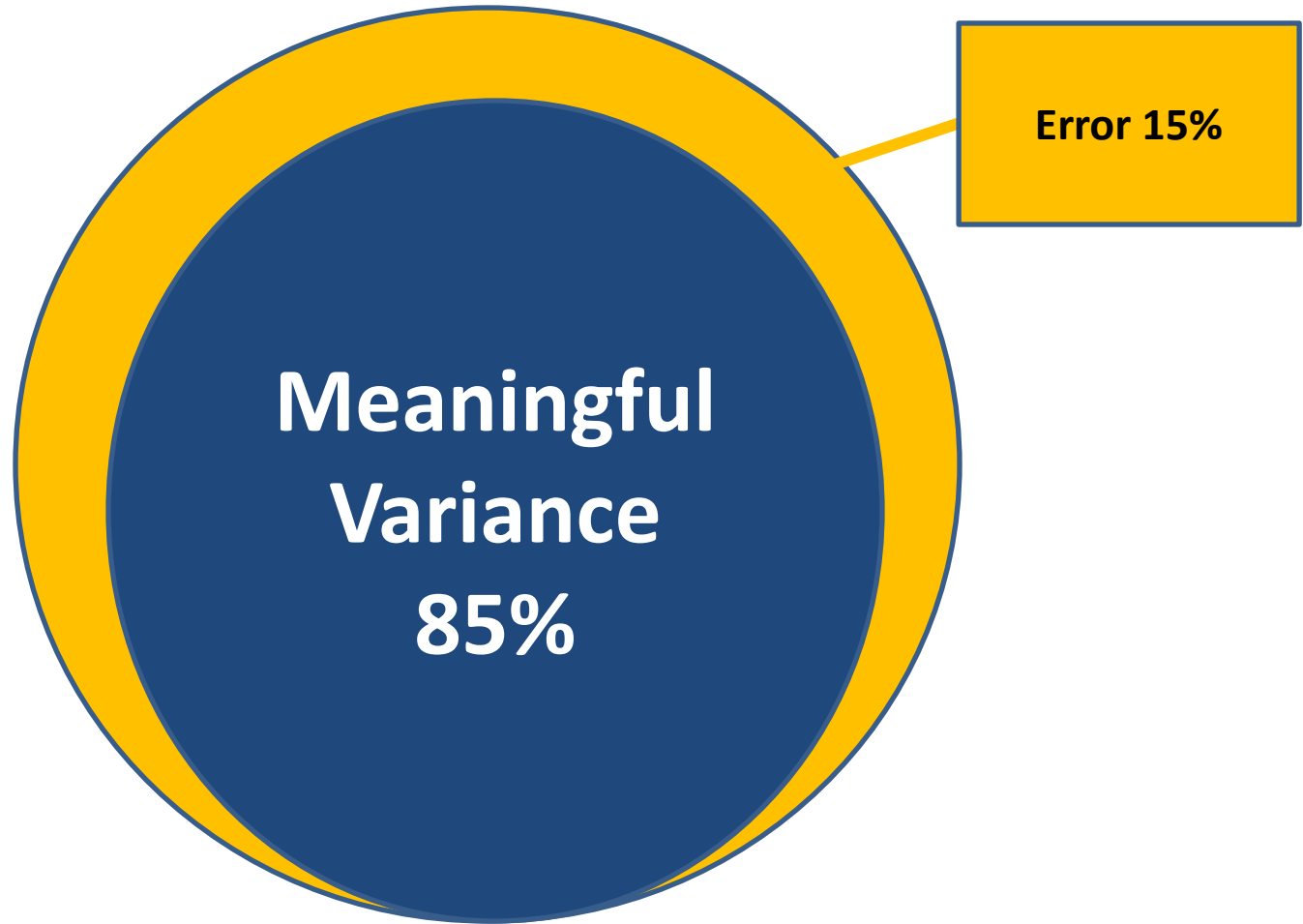
You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# Type

- Test-retest reliability
- Parallel-forms reliability
- Internal consistency reliability (Split-half)

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# Conceptualization of $r_{xx'} = 0.85$



You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# GROUP DISCUSSION

What can be sources of error?

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# Group Work Instructions

(Facilitation Technique: Typology Development Adapted)

1. Each one write three sources of error.
2. Write one source on one piece of post-it provided.
3. Post the completed post-its on the wall.
4. Read what others have written.
5. Group similar ones together.
6. Label the ideas.

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

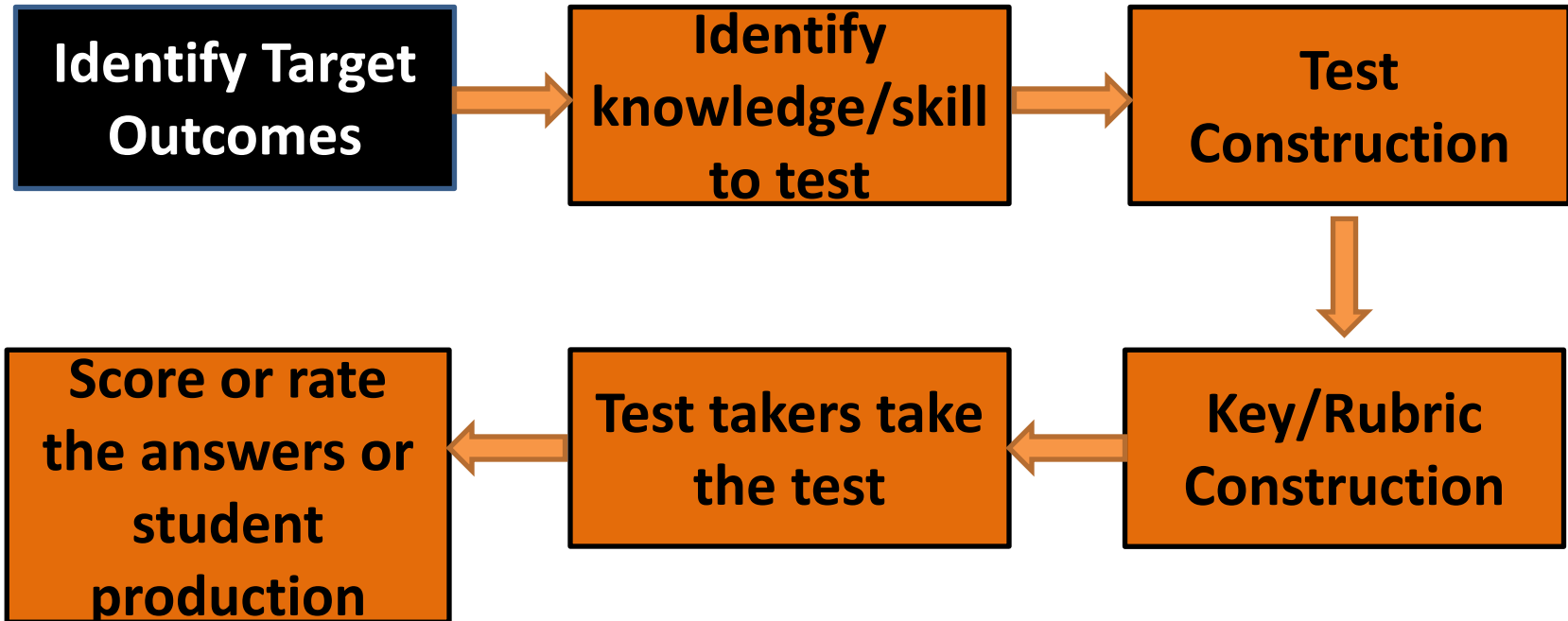
# CHECKLIST FOR POTENTIAL SOURCES OF ERROR

Handout

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.



# Testing Procedure



You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# Ranges of Reliability



You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# Calculation of reliability

- Test-retest reliability – correlation
- Parallel-forms reliability – correlation
- Internal consistency reliability (Split-half) – correlation adjusted for length

$$\text{Spearman-Brown Prophecy } r_{xx'} = \frac{n * r}{(n - 1) * r + 1}$$

- $r_{xx'}$  = *full-test reliability*
- $r$  = *correlation*
- $n$  = *the number of times the test length is to be increased*

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# EXCEL ACTIVITY DEMONSTRATION + HANDS-ON

Please download the reliability calculation Excel practice file.

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# JUDGMENT RELIABILITY

- Interrater agreement
- Interrater reliability
- Intrarater reliability

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

## Rating Results of Student Essays

**0 = Failing**

**1 = Approaching**

**2 = Meeting**

**3 = Exceeding**

StudentID	Rater1	Rater2
101	3	0
102	1	1
103	2	1
...		

$$\text{Interrater Agreement} = \frac{\text{Pairs of same ratings}}{\text{All Pairs}} = 0.33$$

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# Interrater Reliability

- Consistency between raters
- Like parallel forms: The correlation between two raters
- If an aggregated score is used → correlation adjusted for number of raters, like split-half

$$\text{Spearman-Brown Prophecy } r_{xx'} = \frac{n * r}{(n - 1) * r + 1}$$

- $r_{xx'}$  = full-test reliability
- $r$  = correlation
- $n$  = the number of raters

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# Intrarater Reliability

- Like test-retest reliability
- Consistency of judgment over time.
- Correlation
- Be aware of memory effect

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.



# ReCal

- Reliability Calculation for the masses
- URL: <http://dfreelon.org/utlis/recalfront/>

Level of measurement	N of coders	Use
Nominal	2 coders only	<u>ReCal2</u> (includes percent agreement, Scott's pi, Cohen's kappa, and nominal Krippendorff's alpha)
Nominal	3 or more coders	<u>ReCal3</u> (includes pairwise percent agreement, Fleiss' kappa, pairwise Cohen's kappa, and nominal Krippendorff's alpha)
Ordinal, interval, or ratio	Any N of coders	<u>ReCal OIR</u> (includes ordinal, interval, and ratio Krippendorff's alpha)

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# Sample file format

	A	B	C	D	E	F	
1	1	1	1	1	0	0	
2	1	1	0	0	0	0	
3	1	1	0	0	0	0	
4	1	1	0	0	0	0	
5	1	1	0	0	0	0	
6	1	1	1	1	0	0	
7	0	0	0	0	0	0	
8	0	0	0	0	0	0	
9	0	1	0	0	0	0	
10	1	1	0	0	0	0	

R1\_SLO1

R2\_SLO1

R1\_SLO2

R2\_SLO2

Save as comma delimited file

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# HANDS-ON DEMONSTRATION

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

# STRATEGIES TO IMPROVE RELIABILITY

## Group Activity:

1. Install Socrative Student on your tablet or your phone. Or on your computer go to:  
[m.socrative.com/student/#joinRoom](https://m.socrative.com/student/#joinRoom)
2. Answer the short-answer question.

You may use these materials only for nonprofit educational purposes. Please give credit/cite appropriately.

### **Checklist for Sources of Error (Brown, 1996, p. 189)**

- Variation due to environment
  - Location
  - Space
  - Ventilation
  - Noise
  - Lighting
  - Weather
- Variance due to administration procedures
  - Directions
  - Equipment
  - Timing
  - Mechanics
- Variance attributable to examinees
  - Health fatigue
  - Physical characteristics
  - Motivation
  - Emotion
  - Memory
  - Concentration
  - Forgetfulness
  - Impulsiveness
  - Carelessness
  - Test wiseness
  - Comprehension of directions
  - Guessing
  - Task performance speed
  - Chance knowledge of item content
- Variance due to scoring procedures
  - Errors in scoring
  - Subjectivity
  - Evaluator biases
  - Evaluator idiosyncrasies
- Variance attributable to the test and test items
  - Test booklet clarity
  - Answer sheet format
  - Particular sample of items
  - Item types number of items
  - Item quality
  - Test security

Brown, J. D. (1996). *Testing in Language Programs*. Upper Saddle River, NJ: Prentice Hall Regents Prentice-Hall, Inc.

## Typology Development

**Source:** Custer, R. L., & Daugherty, M. K. (2004). Project ProBase and the process to design bridge competencies for community college technical programs. *The Journal of Applied Research in the Community College*, 11(2), 55-64

**Purpose:** to reach consensus of categories and categorical components of ideas/solutions/problems/strategies.

**Time:** 1 - 2 hours for 5 to 10 participants

### Procedure:

1. Give each team member a large, blank Post-it(R) note pad and record as many ideas/solutions/problems/strategies as possible in 3 minutes. Each ideas/solutions/problems/strategies was recorded on a single Post-it(R) note.
2. Ask each participant to pass their note pad to one person on the left. They are directed to review the ideas/solutions/problems/strategies that are passed to them and they are given additional three minutes to add more to the list.
3. Repeat the process until the note pads had rotated through all participants and were returned to the original author.
4. The facilitator will then post each single note on a large white board. Participants are asked to eliminate duplication by placing redundant items on top of one another.
5. After consensus was achieved on these items, participants were then asked to cluster items that they perceived to be conceptually similar by physically moving Post-it(r) notes around the wall until all items have been classified.
6. Elicit consensus to label each cluster.
7. At the end of the session, the facilitator will carefully package the material for later transcription.