Score Reliability

Brian K. Miller, Ph.D.

Introduction

- What is measurement?
  - _____ consideration
  - _____ consideration
- Reliability
- Error
  - _____
  - Non-_____

Reliability

- Degree of:
  1. dependability,
  2. consistency, or
  3. _____
  - …of _____ on a measure

_____ (CTT)

$X_{obtained}$ = _____ + X_{error}

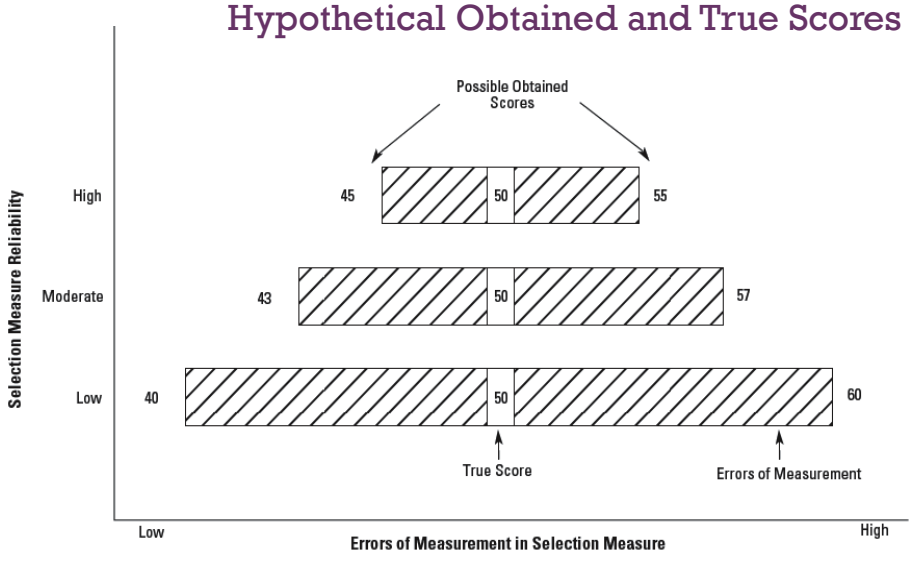
Where:

$X_{obtained}$ = obtained score for a person on a measure

_____ = true score on the measure, that is, actual amount of attribute measured that person really possesses

X_{error} = error score on the measure that is assumed to represent random fluctuations or chance factors

Relationship Between Errors of Measurement and Reliability of a Measure for Hypothetical Obtained and True Scores



Things to Consider Before Selecting Reliability Method

- _____ of assessment?
- Will data hold for future?
- Are scores _____?
- Do raters agree?
- What role does_____play in rating?

Methods of Assessing Reliability

- Test - _____
- Parallel forms
- Internal _____
- Inter-rater

Test-_____ Reliability

- Administer measure _____
- Correlate two sets of scores
- Pearson product-moment _____ coefficient

Illustration of Design for Estimating Test-retest Reliability

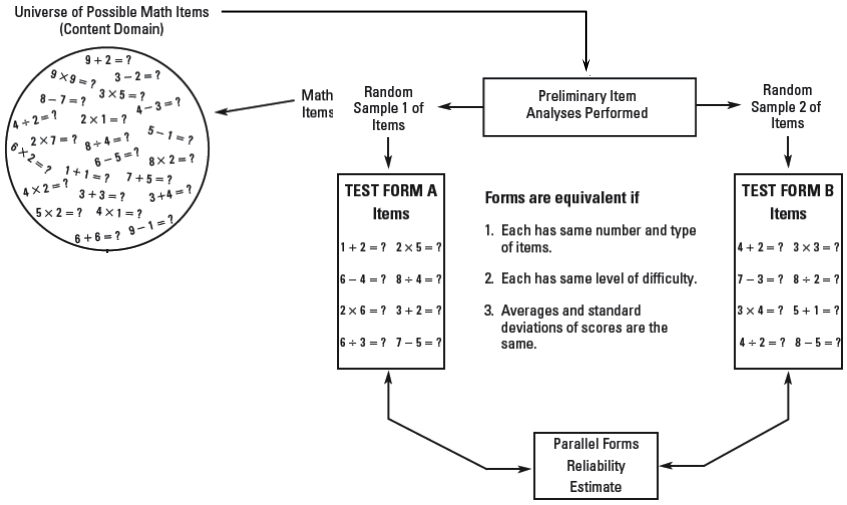
	Test Scores Time 1 (t1)		Retest Scores Time 2 (t2)					
	Time 1	*Rank*	Case A	$t1 - t2_a$	*Rank*	Case B	$t1 - t2_b$	*Rank*
Job Applicant	t1	t1	$t2_a$		$t2_a$	$t2_b$		$t2_b$
J. S. Friedman	**96**	1	**90**	−6	1	**66**	−30	2
J. A. Fukai	**87**	2	**89**	2	2	**52**	−35	3
B. Y. Woodward	**80**	3	**75**	−5	3	**51**	−29	4
T. A. Hinata	**70**	4	**73**	3	4	**82**	12	1
A. C. Zimiski	**56**	5	**66**	10	5	**50**	−6	5

NOTE: Test-retest reliability for Time 1—Time 2 (Case A) = 0.94
Test-retest reliability for Time 1—Time 2 (Case B) = 0.07

Parallel Forms Reliability

- AKA:
  - Equivalent forms reliability
  - _____ forms reliability
- Consistency with which an attribute is measured
- Using _____ of a measure

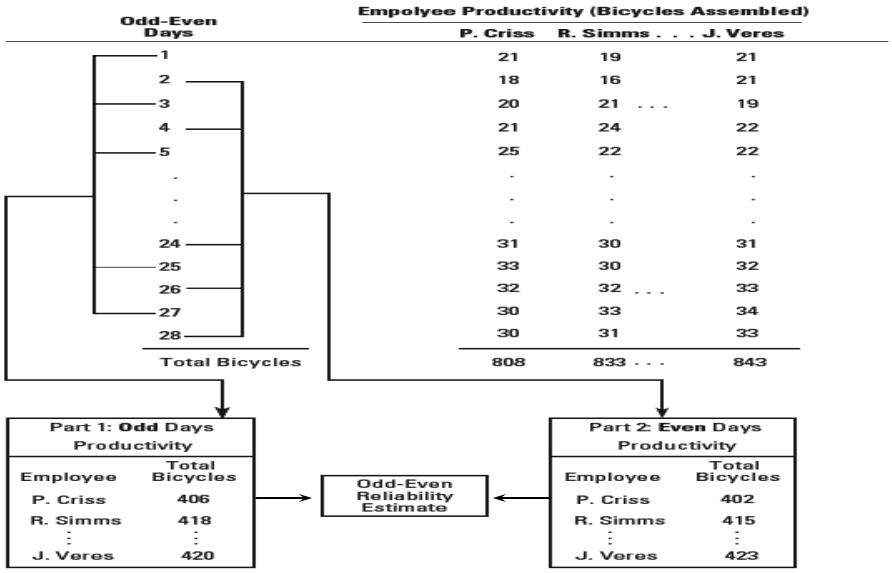
Procedures for Developing Parallel Forms Test



Internal _____ Reliability

- Extent to which all parts of a measure
  - *items or questions*
- Are similar _____
- Variety of methods
  - Split _____
  - KR-20
  - _____ alpha

Development of Odd-even (Day) Split-half Reliability Computed for a Job Performance Criterion Measure



Prophecy Formula

$$r_{ttc} = \frac{nr_{12}}{1 + (n-1)r_{12}}$$

Where:

r_{ttc} = the corrected split-half reliability coefficient for the total selection measure

n = number of times the test was increased in length

r_{12} = the correlation between Parts 1 and 2 of the selection measure

Data Used in Computing K-R 20 Reliability Coefficient

	Test Items								
Applicant	1	2	3	4	5	6	7	8	Test Score
Wiley Boyles	1	1	1	1	1	1	1	1	8
Tammy Allen	0	0	0	1	1	1	0	1	4
Beryl Hesketh	1	0	1	1	1	0	1	1	6
Sidney Craft	0	0	0	0	0	0	1	1	2
David Speed	1	0	1	1	0	0	0	0	3
Cheri Ostroff	1	0	0	1	1	1	1	0	5
Number Correctly Answering Item	4	1	3	5	4	3	4	4	

NOTE: 0 = incorrect response; 1 = correct response.

Formula

$$r_{tt} = \frac{k}{k-1}\left(\frac{\sum p_i(1-p_i)}{\sigma_y^2}\right)$$

Where :

k = number of items on the test

p_i = proportion of examinees getting each item (i) correct

$1-p_i$ = proportion of examinees getting each item (i) incorrect

σ_y^2 = variance of examiner's total test scores

Applicant Data in Computing Coefficient Alpha (α)

	Case 1 Applicants								Case 2 Applicants							
Items Used to Assess Applicant Conscientiousness—A Personality Trait	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Item 1. Dependability	4	1	4	2	4	3	5	2	1	4	4	5	3	2	1	2
Item 2. Organized	5	1	4	2	4	4	5	2	2	3	1	4	3	3	1	5
Item 3. Hardworking	5	1	4	2	5	3	5	3	4	1	3	3	4	4	3	4
Item 4. Persistence	4	1	5	3	5	2	5	3	2	5	5	2	1	5	3	3
Total Score	18	4	17	9	18	12	20	10	9	13	13	14	11	14	8	14

NOTE: Applicants rate their behavior using the following rating scale:
1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, and 5 = Strongly Agree

Case 1 coefficient alpha reliability = 0.83

Case 2 coefficient alpha reliability = 0.40

Cronbach's

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum \sigma_{i^2}}{\sigma_{y^2}}\right)$$

Where :

k = number of items on the selection measure

σ_{i^2} = variance of respondents' scores on each item(i) on the measure

σ_{y^2} = variance of respondents' scores on the measure

Interrater Reliability

- Sources of measurement error
  - _____ is being rated (e.g., employee behavior)
  - _____the rating (rater characteristics)
- Measure of agreement between two or more raters
  - Interrater Agreement
  - _____class Correlation
  - _____class Correlation

Research Design for Computing Intraclass Correlation to Assess Interrater Reliability of Employment Interviewers

	Interviewer		
Interviewee	1	2	3
A. J. Mitra	9	8	8
N. E. Harris	5	6	5
R. C. Davis	4	3	2
•	•	•	•
•	•	•	•
•	•	•	•
T. M. Zuckerman	7	6	7

NOTE: The numbers represent hypothetical ratings of interviewees given by each interviewer.

Factors Influencing the Reliability of a Measure

- _____ of estimating reliability
- Individual differences among respondents
- _____ of a measure
- Test question _____

Factors (cont'd)

- _____ of a measure's content
- Response format
- _____ of a measure

Relation among Test Length, Probability of Measuring True Score, and Test Reliability

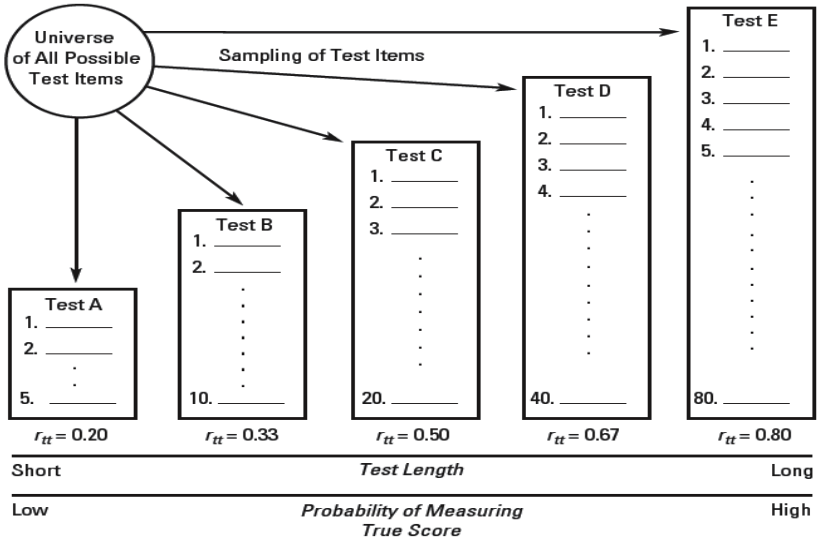
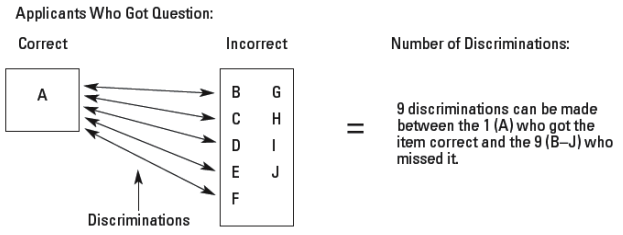


Illustration of Relation Between Test Question Difficulty and Test Discriminability

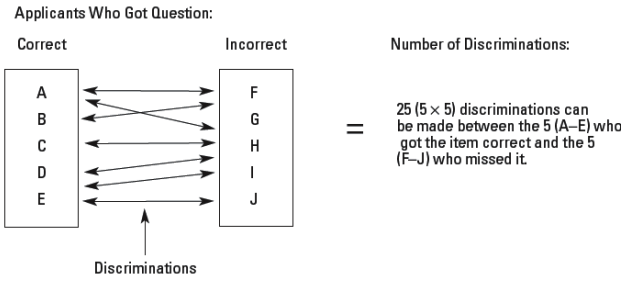
Case 1:
Item Difficulty = 0.10

10% Get Test Question Correct



Case 2:
Item Difficulty = 0.50

50% Get Test Question Correct



Standard Error of _____

- Estimated error in _____ score on the measure
- Calculating standard error of _____

$$\sigma_{meas} = \sigma_x\sqrt{1 - r_{xx}}$$

where

σ_{meas} = the standard error of measurement for measure X

σ_x = the standard deviation of obtained scores on measure X

r_{xx} = the reliability of measure X.

Uses of the SEM

1. Shows that scores are approximation represented by _____ of scores on measure
2. Aids decision making in which only _____ _____ is involved
3. Can determine whether scores for individuals _____ from one another
4. Helps establish confidence in scores obtained from different groups of respondents

Guidelines for Interpreting Individual Scores Using SEM

- Difference between two individuals' scores should not be considered _____unless difference is at least twice the SEM of measure
- Before difference between scores of same individual on two different measures should be treated as significantly different, the difference should be greater than _____ _____of either measure

That's all folks!

BrianMillerPhD@gmail.com