

*Survey Methods & Design in
Psychology*

Lecture 6
Reliabilities, Composite Scores &
Review of Lectures 1 to 6 (2007)

Lecturer: James Neill

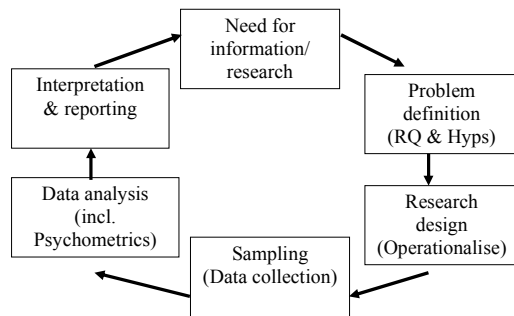
Overview

- Concepts & their measurement
- Psychometrics
- Reliability
- Validity
- Composite scores
- Writing up a factor analysis
- Review of W1-W5
 - Student questions

Readings

- Francis – Ch6.1 (Reliability)
- pp.63-65 “Concepts & their Measurement”
(e-reserve)
- DeCoster, J. (2005). Scale construction notes.
<http://www.stat-help.com/notes.html>

*An Iterative Model of
The Scientific Research Process*



Concepts & Their Measurement

Concepts – Bryman & Cramer

- **Concepts** form a linchpin in the process of social research
- **Concepts** express common elements in the world to which we give a name
- Hypotheses contain **concepts** which are the products of our reflections on the world.

*An Iterative Model of
The Scientific Research Process*

Once formulated, a concept and the concepts with which it is purportedly associated, such as social class and authoritarianism, will need to be *operationally defined*, in order for systematic research to be conducted in relation to it. An operational definition specifies the procedures (operations) that will permit differences between individuals in respect of the concept(s) concerned to be precisely specified. What we are in reality talking about here is *measurement*, that is, the assignment of numbers to the units of analysis – be they people, organizations, or nations – to which a concept refers.

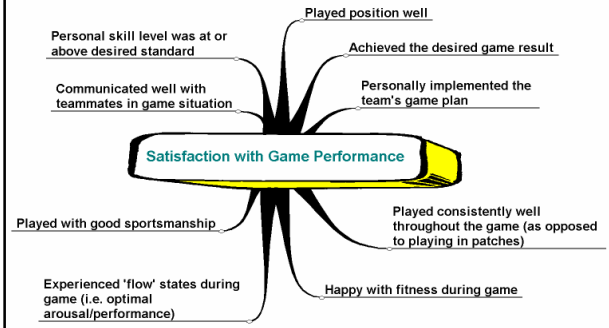
Measuring Concepts - Process

- Brainstorm indicators of concept
- Define concept
- Operationalise – draft a measurement device
- Pilot test
- Examine psychometric properties
– how *precise* are the measures?
- Redraft/refine and re-test

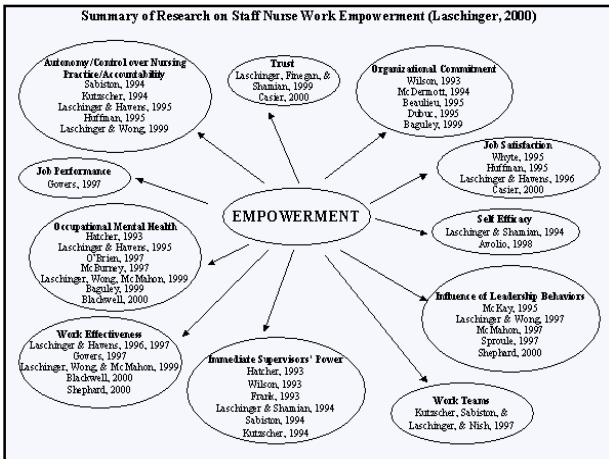
Operationalisation

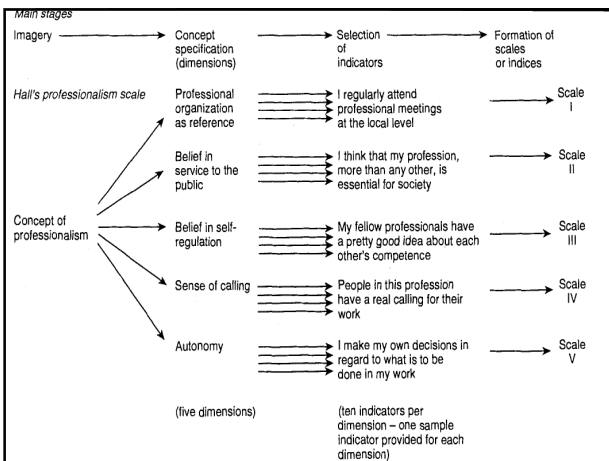
- Operationalisation is the act of making a fuzzy concept measurable.
- Social sciences often use multi-item measures to assess related but distinct aspects of a fuzzy concept.

Fuzzy Concepts - Mindmap



Summary of Research on Staff Nurse Work Empowerment (Laschinger, 2000)





Psychometrics - Goal

Goal:

- To validly measure differences b/w individuals and groups in psychosocial qualities such as ability, attitudes, and personality.

Psychometrics - Tasks

Tasks:

- The construction of instruments and procedures for measurement; and
- The development and refinement of theoretical approaches to measurement.

Psychometric Methods

- Exploratory Factor Analysis
- Classical Test Theory
 - Reliability
 - Validity

*As Test-Taking Grows, Test-Makers
Grow Rarer*

"Psychometrics, one of the most
obscure, esoteric and cerebral
professions in America, is now
also one of the hottest."

- May 5, 2006, *NY Times*.

**So You Want to Be
A Psychometrician?**

Becoming a psychometrician, or
testing expert, requires years of
graduate work.

COURSE WORK AT U. OF IOWA
Ph.D. in Educational Measurement
and Statistics

Applied Statistics

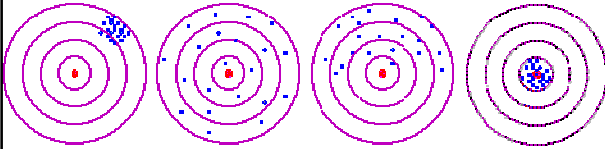
- Intermediate Statistical Methods
- Correlation and Regression
- Design of Experiments
- Nonparametric Statistical Methods
- Factor Analysis and Structural Equation Models
- Introduction to Multivariate Statistical Methods
- Topics in Educational Measurement and Statistics

Educational Measurement

- Construction and Use of Evaluation Instruments
- Educational Measurement and Evaluation Theory and Technique in Educational Measurement
- Scaling Methods
- Item Response Theory
- Seminar: Educational Measurement and Evaluation
- Equating and Scaling of Educational Tests
- Generalizability Theory
- Appraisal in Counseling

Reliability & Validity

Reliability vs Validity



**Reliable
Not Valid**

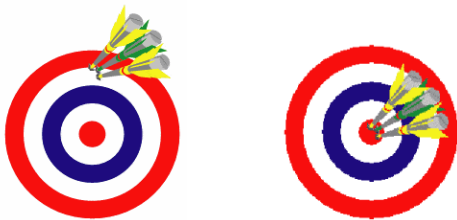
**Valid
Not Reliable**

**Neither Reliable
Nor Valid**

**Both Reliable
And Valid**

Reliability

Reproducibility of a measurement



Types of Reliability

- **Internal consistency**
 - correlations amongst multiple items in a factor
 - e.g., Cronbach's Alpha (α)
- **Test-retest reliability**
 - correlation between time 1 & time 2
 - e.g., Product-moment correlation (r)

Reliability Interpretation

- <.6 = not reliable
- .6 = OK
- .7 = reasonably reliable
- .8 = good, strong reliability
- .9 = excellent, very reliable
- >.9 = potentially overly reliable or redundant measurement – this is subjective and whether a scale is overly reliable depends also on the nature what is being measured

Variable	<i>Journal of Personality and Social Psychology</i>		<i>Journal of Applied Psychology</i>	
	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Average reliability of variables				
Less than .60	3	1.9	2	3.4
.60-.69	6	3.8	5	8.6
.70-.79	33	20.8	9	15.5
.80-.89	33	20.8	11	19.0
.90-1.00	14	8.8	9	15.5
Unknown	70	44.0	22	37.9

Internal Reliability

- Is a multi-item scale measuring a single concept?
- Are items in scale consistent with one another?

Types of Internal Reliability

- Split-half reliability
The first half of the items are summed and then correlated with the sum of the second half of the items.
- Odd-even reliability
Items 1, 3, 5, etc. are summed and correlated with Items 2, 4, 6, etc..
- Alpha reliability (Cronbach's α)
Averages all possible split-half reliability coefficients - akin to a single score representing the extent of intercorrelation amongst the items

How Many Items per Factor?

- More items -> greater reliability
(The more items, the more 'rounded' the measure)
- Law of diminishing returns
- Min. = 3
- Max. = unlimited
- Typically ~ 4 to 10 is reasonable
- Final decision is subjective and depends on research context

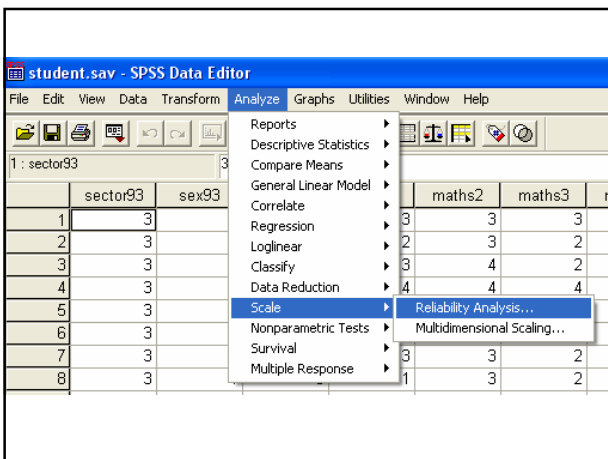
Internal Reliability – Quality of Maths Class Example

- 10-item scale measuring students' assessment of their maths classes
- 4-point Likert scale from: strongly disagree to strongly agree.
- Ensure -ve items are recoded

Quality of Mathematics Teaching

1. My maths teacher is friendly and cares about me
2. The work we do in our maths class is well organised.
3. My maths teacher expects high standards of work from everyone.
4. My maths teacher helps me to learn.
5. I enjoy the work I do in maths classes.

+ 5 more



SPSS - Corrected Item-Total Correlation

Reliability Statistics

Cronbach's Alpha	N of Items
.885	10

A measure for examining the relationship between individual items and the total scale, this is the correlation between the given item and the item sum if the given item is not included in the scale. Smaller values indicate the given item is not well correlated with the others.

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
school1	41.15	98.608	.438	.888
school2	40.04	91.500	.648	.872

SPSS - Cronbach's Alpha

Reliability Statistics

Cronbach's Alpha	N of Items
.885	10

A measure for examining the relationship between individual items and the total scale, this is the value of Cronbach's Alpha for the remaining items if the given item is not included in the scale.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
school1	41.15	98.608	.438	.888
school2	40.04	91.500	.648	.872

MATHS3 25.0192 30.5174 .0996 .9021

Alpha = .8790

Alpha = .9024

Internal Reliabilities for Classroom Behaviour

Behav.sav example:

- Factor 1 (Attentiveness) $\alpha = .94$
- Factor 2 (Settledness) $\alpha = .89$
- Factor 3 (Sociability) $\alpha = .90$

Table *. Definitions of the Life Effectiveness Questionnaire dimensions, with Internal Consistency and Test-Retest Correlations

LEQ 8-factor model	Description	3 items per scale	
		Test-Retest <i>r</i>	Alpha
Achievement Motivation	Motivation to achieve excellence and put the required effort into action to attain it.	.68	.87
Active Initiative *	Initiating action in new situations.	.73	.81
Emotional Control	Maintaining emotional control when faced with potentially stressful situations.	.75	.87
Intellectual Flexibility	Adapting thinking and accommodating new information from changing conditions and different perspectives.	.60	.78
Self Confidence *	Confidence in abilities and the success of actions.	.73	.84
Social Competence	Ability in and success of social interactions.	.75	.86
Task Leadership	Ability to lead other people effectively when a task needs to be done and productivity is the primary requirement.	.81	.82
Time Management	Makes optimum use of time.	.75	.84
Total	Effective in generic life skills.	.72	.84
<i>N</i>		.67	.93

(Construct) Validity

To extent to which an instrument actually measures what it purports to measure.



Types of Validity

- Construct validity
 - Translation validity
 - Face validity
 - Content validity
 - Criterion-related validity
 - Predictive validity
 - Concurrent validity
 - Convergent validity
 - Discriminant validity

Types of Validity – “Translation”

Face validity

- Prima facie extent to which an item is judged to reflect target construct

Content validity

- Systematic examination of the extent to which test content covers a representative sample of the domain to be measured – e.g. sources,
 - existing literature
 - expert panels
 - qualitative interviews / focus groups with target sample

Types of Validity – “Criterion”

Concurrent validity

- Correlation between the measure and other recognised measures of the target construct

Predictive validity

- Extent to which a measure predicts something that it theoretically should be able to predict.

Types of Validity – “Criterion”

Convergent validity

- Extent to which a measure correlates with measures with which it theoretically should be associated.

Discriminant validity

- Extent to which a measure does not correlate with measures with which it theoretically should not be associated.

Composite Scores (Factor Scores)

Used to reliably estimate individual differences in target constructs.

Univariate, continuous-like variables which can be used for:

- Descriptives, screening, testing, feedback
- As IVs or DVs in subsequent inferential analyses

Creating Composite Scores

Two methods:

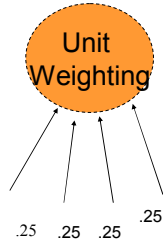
- Unit weighting
- Factor score regression weights

Unit Weighting

Unit Weighting

Average or total of all variables in a factor
(i.e., each variable is equally weighted).

$$\bar{X} = \text{mean}(y_1 \dots y_p)$$



Creating Composite Scores – Dealing with Missing Data

It can be helpful to maximize sample size by
estimating some of the missing values.

Composite Scores – Missing Data

A technique in SPSS for the calculation of
composite factor scores which allows for some
missing items:

$$\bar{X} = \text{mean} (v_1, v_2, v_3, v_4, v_5, v_6)$$

$$\bar{X} = \text{mean}.4 (v_1, v_2, v_3, v_4, v_5, v_6)$$

Creating Composite Scores – Dealing with Missing Data

How many items is it OK to allow to be missing? A guide:

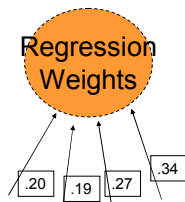
- 1 item missing OK per 4 to 5 items
- 2 items missing OK per 6 to 8 items
- 3+ items to be missing for 9+ items

Regression Weighting

Factor Score Regression Weights

The contribution of each variable to the total is weighted to reflect some items more than other items.

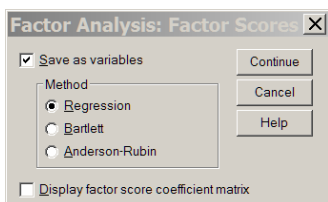
$$X = 20*v_1 + .19*v_2 + .27*v_3 + .34*v_4$$



Regression Weighting

Two calculation methods:

- Manual (use Compute)
- Automatic (use Factor Analysis – Factor Scores)



Writing up a factor analysis

- Discussion
 - Theoretical underpinning – Supported? Adaptations?
 - Quality / usefulness of measure
 - Recommendations for further improvement
- More information
 - Writing up a factor analysis (.doc)
 - <http://wilderdom.com/courses/surveyresearch/assessment/labreport/>

Review: Q & A of Week 1 to 6

Add questions into the box.

References

Howitt, D., & Cramer, D. (2005). Chapter 13
Reliability and validity: Evaluating the value of
tests and measures. In *Introduction to research
methods in psychology* (pp. 218-231). Essex, UK:
Pearson.

Presser, S., Rothgeb, J. J., Couper, M. P., Lessler, J.
T., Martin, E., Martin, J., & Singer, E. (Eds.)
(2004). *Methods for testing and evaluating survey
questionnaires*. Wiley: Hoboken, NJ.
