Measurement Theory

Introduction to Validity and Reliability

1

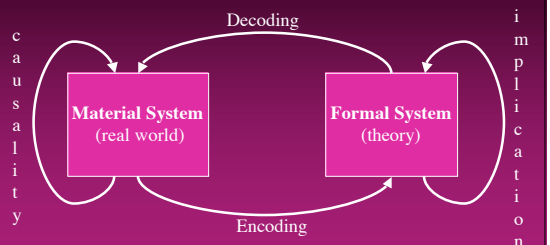
Measurement Theory

2

I. Newton's scheme for mathematical modeling
II. Operationism
III. Validity of Measurement
IV. Internal versus External Validity
V. Reliability of Measurement
VI. Conclusions

I. Newton's scheme for mathematical modeling

3



Example: Predicting the motion of the planets

4

II. Operationism

In Newton's system, how we encode and decode.

Forming operational definitions of our concepts.

A. **Independent variable**: the variable that the experimenter directly changes.

- must be defined in universal terms
- must be something the experimenter can directly control

5

II. Operationism (cont)

B. **Dependent Variable**: The variable(s) measured or examined.

In an experiment: one looks for changes in the dependent variable as a result of changes in the independent variable

In observational (correlational) research, one looks for associations between dependent variables.

6

Dependent variable (cont)

- must be defined in universal terms

- must be readily observable

7

II. Operationism (cont)

C. Two primary issues in defining independent and dependent variables.

Validity: the accuracy of the definition

Reliability: the consistency of the measurement techniques

8

Validity vs. Reliability

Example: bathroom scales

Validity: accuracy of the scale
does it measure what you want
(pounds vs. fat or dress size)

Reliability: does the scale give consistent results?

9

III. Validity of Measurement

A. **Content validity**

Does the test or measurement device (i.e. the dependent variable) provide a direct and representative measure of the concept?

Example: praise and school grades

Is our measurement of grades direct and representative?

10

III. Validity of Measurement

B. **Criterion validity**:

Is the measure associated with other related measures? Does it match a standard or criterion measure?

Example: If we are trying to get an index of depression, how does our measure compare to the Beck Depression Inventory?

11

III. Validity of Measurement

C. **Construct Validity**

Does the measurement device actually measure the theoretical construct that we say it measures?

Converging Operations?

Example: Use of student ratings as a measure of teacher effectiveness.

Is this a measure of teaching skill or a measure of how much the students like the teacher?

12

IV. Internal versus External Validity

A. Internal Validity:

Are the causal inferences or the empirical conclusions accurate?

Internal validity requires that alternative explanations of the results have been eliminated.

IV. Internal versus External Validity

13

B. External Validity

Are the results of the research generalizable to other situations? Do they apply to the settings of primary interest?

External validity requires good construct and content validity.

IV. Internal versus External Validity

14

C. Often there is a trade-off between internal and external validity.

V. Reliability of Measurement

15

A. Experimental Reliability:

1) Replication: repeating an experiment or observation should lead to the same results.

2) Statistical Reliability: based on the sample data, we should be able to expect the same results from a new sample.

16

V. Reliability of Measurement

B. Test Reliability

1) test-retest: If we administer the same test a second time, we should get the same results.
2) parallel forms: Different versions of the same test should yield the same results.
3) split-half: Different sections of the test should yield the same results.

17

VI. Conclusions

1. Research is only as good as the definitions it uses (how concepts are decoded and encoding)
2. Definitions must be operational
3. Definitions must be valid.
4. Measurement must be reliable.

18

Discussion

Imagine you are trying to make an important decision. You decide to try to make that decision in a rational, scientific manner.

Which is worse, no research or bad research?