Introduction to Test Theory & Historical Perspectives

Measurement Methods in Psychological Research

Lecture 2 – 01/31/2006

Today's Lecture

- General introduction to test theory/what we will cover in this course.
- Historical perspectives in psychometrics.

Theory of Measurement = Test Theory

- Test theory is an abbreviated expression for:
  - Theory of Psychological Tests and Measurements.
- Test theory provides a general collection of techniques for evaluating the development and use, in assessment, of specific psychological tests.

What Test Theory Is Not

- Test theory is *not* the same as assessment.
  - Assessment is the evaluation of characteristics of individuals through use of tests (or other mechanisms).
- Test theory is narrower more general in scope.
  - Just as experimental design is more general and narrower than any survey of actual experimental studies in an area of psychology.

Practical Problems in Measurement

- To demonstrate the types of issues we will discuss, consider the following three examples of measurement:
  1. A teacher wishing to evaluate student knowledge of a math unit.
  2. A clinical psychologist wishing to develop a pencil-and-paper based measure of a psychological disorder.
  3. A survey researcher wishing to study attitudes toward gun control.

Example #1 – The Math Teacher

- A teacher constructs 20 items for a math test.
- Each item can be scored either pass or fail.
- The teacher gives the test and adds up the number of passes to make a math score for each student.
- In doing so, the teacher wonders…

Questions…

- Should there be one score for math or two scores?
  - A score for geometry items and a score for algebra items
- What about items that require both algebra and geometry.
- Are all the items good measures of math ability or are some better than others?

More Questions…

- If one score is sufficient, how accurate is it as a measure of math ability?
- Are 20 items sufficient to give a reasonably accurate determination of each student's knowledge
  - Should more be used?
  - Could fewer have been used?

Even More Questions…

- If different items had been written, would they have measured the same thing?
  - Equally well?
  - Can two tests be made (with different items) so that the scores are interchangeable?
  - Could a computer be used to administer the test adaptively?

Still Even More Questions…

- Are students who have low scores measured as accurately as students scoring highly or in the middle?
- Are the items free from bias when given to students of different cultural backgrounds?
  - Could some students have irrelevant problems with certain items because of differences in their background and experience?
  - How would we be able to know?

Example #2 – The Clinical Psychologist

- A clinical psychologist writes a set true/false of items such as:
  - I have difficulty sleeping.
  - I am afraid of heights.
  - I get tired easily.
  - I often have bad dreams.

- The clinical psychologist wonders the same things as the teacher does…

Example #3 – The Survey Researcher

- A survey researcher wants to study attitudes to gun control.
- A series of items are created that read:
  - Assault weapons do not belong in private hands.
  - All hand guns should be licensed.
  - Government interference with the right to bear arms is an infringement of liberties.

Test Theory

- Test theory consists of the use of mathematical concepts that have been developed in order to refine questions such as these into more precise forms and to provide answers to them.
- Test theory is essentially applied mathematics, overlapping with statistics.

Test Theory

- The development of test theory has been motivated in large part by problems posed in psychology and educational psychology.
- Historically, psychologists have been the primarily developers of the field.
  - Although many struggled with the mathematics.
- In recent decades, mathematicians have recognized the problem, and have added to the field.

Early Days

- Advances in test theory were hampered in the early days by lack of computing.
- Many methods created before the 1950s were rough approximations, created to limit computing time.
  - Our author, Rod, calls things like rough approximations “devices.”

Development of Test Theory

- Begins about the mid 19th century in psychophysical laboratories.
- 1904: Charles Spearman published two seminal papers:
  1. One showed how to recognize from test data that the tests measure just one psychological attribute in common, which led to development of common factor theory.
  2. The other one showed how to estimate the amount of error in test scores, which gave rise to classical true score theory.

Common Factor Theory: 1900s

- Spearman's work:
  - Cognitive performances depend on a unitary psychological function (general intelligence).
  - Motivated developments of intelligence tests:
    - Stanford-Binet.
    - Wechsler.

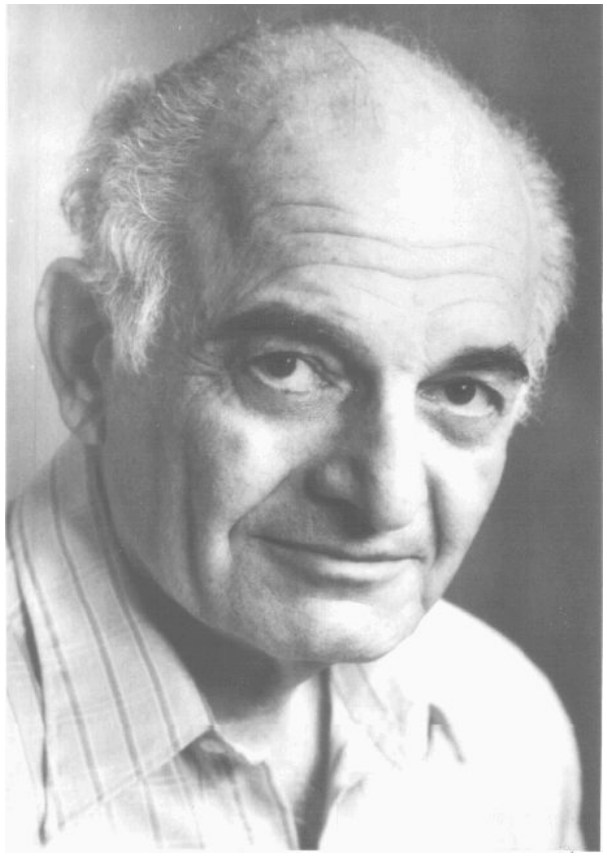
Common Factor Theory: 1930s

- Thurstone elaborated Spearman's model into "multiple factor" model.
  - Later applied in personality tests.

L. L Thurstone

Common Factor Theory: 1940s

- Louis Guttman showed that factor analysis and test development was essentially about generalizing from measures we have created to more measures of the same kind.



Louis Guttman

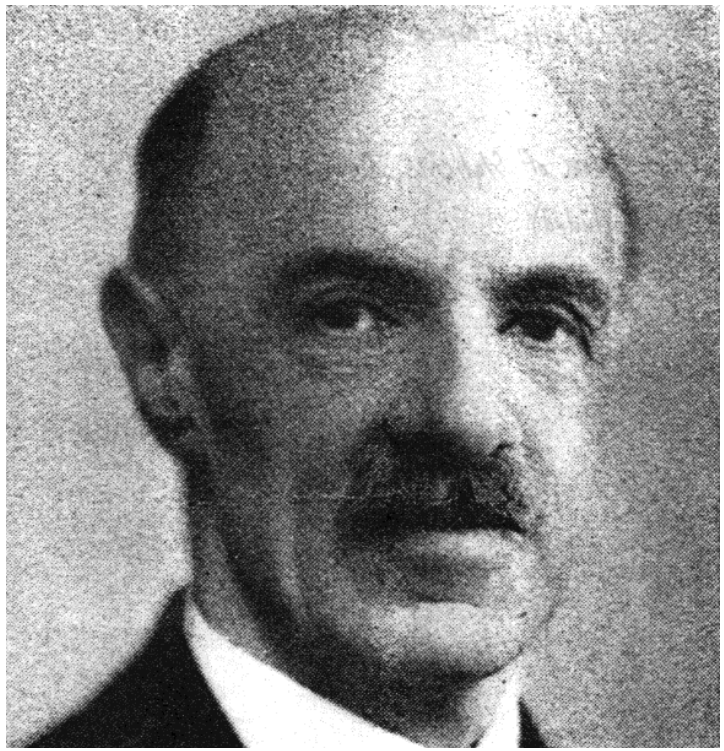
Common Factor Theory: 1940s

- Lawley gave a rigorous statistical treatment of common factor analysis.
- Lawley further gave confirmatory form of factor model.
  - This was later extended by Howe and Bargamann and later, Jöreskog.

Karl Jöreskog

Classical Test Theory

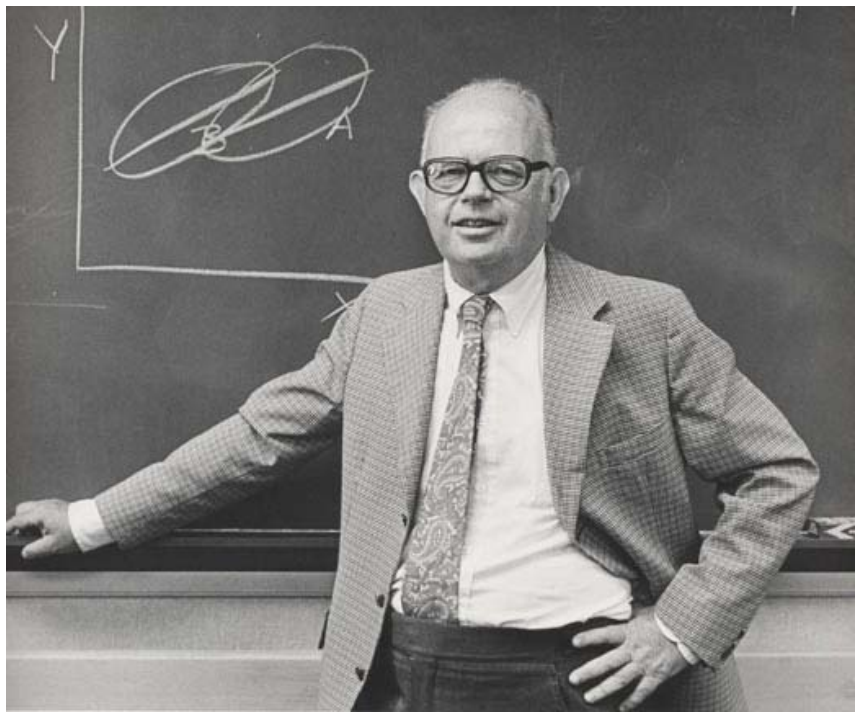
- 1904: Spearman showed that test reliability can be obtained from correlation between two alternative form or two sets of scores from the same test.



Charles Spearman

Classical Test Theory: 1940s – 1950s

- 1945: Guttman showed reliability could be obtained from the relations between the items of a test.
- 1951: Cronbach further development on Guttman's work led to "Cronbach's alpha," which further elaborated into generalizability theory.



Lee Cronbach

Item Response Theory

- IRT is the combination of factor analysis and phi-gamma laws.
  - From 19th century psychophysiology.
- 1950: Lazarsfeld introduced latent structure theory which extended factor analytic method to binary responses.
  - Becomes part of IRT (aka, item factor analysis or latent trait theory).

Paul Lazarsfeld

Item Response Theory

- 1952: Frederic Lord demonstrated a form of Spearman's single factor model can be applied to binary item scores and later elaborated further by Birnbaum in 1960s.

Unified Treatment of Test Theory

- Connection between common factor theory, classical test theory, and item response theory.
- Common factor model can be used as linear approximation to the item response model when applied to binary data.
- Main proponent of this approach: Roderick McDonald (1999)

Wrapping Up…

- As Rod says:
  - Psychology (and test theory) has a long past but a short history.
- We will cover all three main topics in test theory:
  - Classical Test Theory
  - Common Factor Model
  - Item Response Theory
- All topics will be covered from a unified framework (with the common factor model).

Next Time

- General item types used in many tests and survey instruments.
- Item scoring methods.
- How these relate to what we are about to cover in class.