The Importance of Psychological Measurement

PSYC3302: Psychological Measurement and Its Applications

Mark Hurlstone
University of Western Australia

Week 1
Learning Objectives

- Importance of psychological measurement
- Measuring unobserved psychological attributes by observing behaviour
- Psychological tests
  - definition
  - types of tests
- Importance of individual differences in psychological measurement
This unit is about *psychometrics*

Psychometrics is the science of measuring psychological abilities, attributes, and characteristics.

These abilities, attributes, and characteristics are measured using psychological tests.

Psychometrics is concerned with the mathematical, statistical, and professional protocols that underpin tests—how tests are constructed, used, and evaluated.
This unit is **not** about *psychological assessment*

Psychological assessment is concerned with the different types of tests that have been published, how to administer them, and how to interpret the scores.

Psychometrics is concerned with evaluating the *attributes* of psychological tests:

1. the types of information (usually test scores) generated by tests
2. the reliability of data from tests
3. the validity of data from tests

An understanding of psychometrics is of greater utility—it will allow you to construct and evaluate tests of your own, as well as those of others.
Psychological measurement is pervasive—we have all taken psychological tests at some point in our lives.

Everyday, across the world, psychological tests are used to make important decisions:

- *what is the patient’s diagnosis?*
- *is this person competent to stand trial?*
- *who should be hired, promoted, or fired?*
- *which student should be awarded a scholarship?*
- *which parent should gain custody?*

Sometimes, these psychological tests are used to make life or death decisions.
US state of Oregon has an aid-in-dying law—the Oregon Death With Dignity Act (ODDA, 1997)

Provides a patient with 6-months or less to live with the provision to end her or his life through a lethal medical dose

Assistance will be denied to individuals “suffering from a psychiatric or psychological disorder, or depression causing impaired judgement” (ODDA, 1997)

Requires a psychological evaluation of the patient by a state licensed psychologist

Assessment of psychopathology is made on the basis of formal psychological tests of depression, anxiety, dementia, delirium, and psychosis
• In the US state of North Carolina mentally retarded prisoners cannot be subjected to the death penalty

• "A mentally retarded person convicted of first degree murder shall not be sentenced to death" (Criminal Procedure Act, 2007)

• The Criminal Procedure Act defines mental retardation as general intellectual functioning that is "significantly sub-average"

• "significantly sub-average" is a score of 70 or less on an "individually administered, scientifically recognized, standardized intelligence quotient test administered by a licensed psychiatrist or psychologist"
Responses on psychological tests can therefore have a significant impact on people’s lives. Accordingly, psychological test developers and assessors must have confidence in the tests that they employ. They need to know what does and does not constitute a "good test"—this requires knowledge of fundamental principles of psychological measurement. Without this knowledge, they run the risk of harming test-users. In this unit, you are going to learn about the key attributes that make a good test.
Psychological Measurement: Reliability & Validity

• Psychometricians often speak of the *psychometric soundness* of tests

• There are two key aspects of psychometric soundness: *reliability* and *validity*

• The reliability of a measuring tool refers to its consistency—the precision with which the test measures, and the extent to which error is present in the measurements

• The validity of a measuring tool, for a particular purpose, refers to the degree to which it measures what it purports to measure

**Note:**

• We will be discussing reliability and validity in much more detail in subsequent lectures
Psychometricians often speak of the *psychometric soundness* of tests.

- There are two key aspects of psychometric soundness: *reliability* and *validity*.

- The reliability of a measuring tool refers to its consistency—the precision with which the test measures, and the extent to which error is present in the measurements.

- The validity of a measuring tool, for a particular purpose, refers to the degree to which it measures what it purports to measure.

**Note:**

- We will be discussing reliability and validity in much more detail in subsequent lectures.
Unobservable Psychological Attributes

- When it comes to measurement, psychologists do not have it easy.
- Researchers in the natural sciences (chemistry, medicine, physics) can (usually) directly observe the objects of their study.
- The objects of study in psychology are psychological attributes (e.g., working memory, intelligence, executive functioning) which are not directly observable.
- Instead, psychologists typically observe human behaviour to draw inferences about these unobservable psychological attributes.
Unobservable Psychological Attributes

- To do so, we first identify some aspect of behaviour that we think reflects the unobservable psychological attribute of interest.
- We then measure that behaviour via some instrument and interpret the measurements in light of the unobservable psychological construct believed to be reflected in the behaviour.
- In most cases, we develop psychological tests to sample behaviour that we think is sensitive to the hypothesised psychological attributes.
- Let’s consider a concrete example ...
Example: Measuring Working Memory

- Working memory is a cognitive system that supports general cognition by *processing* and *storing* information.
- A number of tasks have been developed that tap this competency, one of which is *Operation Span* (Turner & Engle, 1989).
- In this task, participants are required to solve a series of math operations while trying to remember the order of a set of unrelated words.
- Word–operation strings are presented in sets of two to five items.
- There are three trials for each set size, presented in random order to the participant.
Example: Measuring Working Memory

Example of a three-item set:

IS (8 / 2) – 1 = 1? BEAR
Example: Measuring Working Memory

Example of a three-item set:

\[
\text{IS } (6 \times 1) + 2 = 8? \text{ DRILL}
\]
Example: Measuring Working Memory

Example of a three-item set:

IS \((10 \times 2) - 5 = 15\) JOB
Example: Measuring Working Memory

**Example of a three-item set:**

???
Example: Measuring Working Memory

Example of a three-item set:

BEAR
Example of a three-item set:

BEAR DRILL
Example: Measuring Working Memory

Example of a three-item set:

BEAR DRILL JOB
Example: Measuring Working Memory

- To ensure participants do not trade off solving operations and remembering words, they must maintain 85% accuracy on the math operations.

- A person’s operation span is the total number of words recalled in the correct order across all trials and set sizes (a value between 0 and 54).
Example: Measuring Working Memory

- A person’s operation span is taken to be a measure of their working memory.
- Thus, if one person (person A) has an operation span of 44, and another (person B) has an operation span of 28, we would conclude that person A has a larger working memory than person B.
- This conclusion requires that we make an *inference*—that an overt behaviour (a person’s operation span) is systematically related to the psychological attribute that is working memory.
- For this interpretation to be valid, the operation span task must be theoretically linked to working memory.
Example: Measuring Working Memory

- Recall from the earlier slide that working memory is defined as a cognitive system that combines short-term *storage* and *processing*.

- On the face of it, the operation span task therefore appears to be a valid measure of working memory since it combines short-term storage (memory for words) with a processing element (solving math operations).

- Indeed, validity studies have shown that operation span has good *construct validity* (more on this in future lectures).

- Measurement in psychology typically—but not always— involves some type of theory linking psychological attributes to an observable behaviour.
Unobservable Psychological Attributes: Latent Variables & Constructs

- Up until now, we have referred to the unobserved theoretical characteristics as psychological attributes.
- These unobserved theoretical characteristics are more commonly referred to as *hypothetical constructs* or *latent variables*.
- They include such things as learning, intelligence, motivation, attitudes, and feelings.
- The operations or procedures used to measure these constructs are known as *operational definitions*.
Psychological Tests

- Psychological tests are the vehicles by which psychologists study hypothetical psychological constructs

- But what exactly is a psychological test?
Psychological Tests

- A psychological test "is a systematic procedure for comparing the behavior of two or more people" (Cronbach, 1960, p.21)

- The definition has three important components:
  1. Tests involve behavioural samples
  2. The behavioural samples are collected in a systematic way
  3. The purpose of the test is to compare the behaviour of two or more people
The Beck Depression Inventory (BDI; Beck et al., 1996) is a 21-item multiple choice test designed to measure depression.

**Example item:**
(0) I do not feel sad.
(1) I feel sad.
(2) I am sad all the time and I can’t snap out of it.
(3) I am so sad or unhappy that I can’t stand it.

The BDI satisfies Cronbach’s criteria for a test:

1. The responses to the 21-items are behavioural samples
2. The test is systematic because everyone completes it in the same way under the same conditions
3. The test can be used to compare two or more people on the dimension of depression
Psychological Tests: Example

- The Operation Span task described previously also meets Cronbach’s criteria for a test
- The responses on the test—viz. words recalled—constitute behavioural samples (Cronbach’s first criterion)
- The task employs a systematic testing procedure to compare the performance of two or more people on the dimension of working memory (Cronbach’s second and third criteria)
Psychological Tests: Types

- There are a wealth of tests that have been published.
- They tend to vary along various different dimensions:
  - content (e.g., skill, personality, attitudes)
  - types of responses (e.g., multiple-choice, open-ended)
  - administration procedure (e.g., individual vs. group)
  - time constraints (speed vs. power)
  - intended purpose (criterion vs. norm referenced)
Psychological Tests: Types

- There are a wealth of tests that have been published
- They tend to vary along various different dimensions:
  - content (e.g., skill, personality, attitudes)
  - types of responses (e.g., multiple-choice, open-ended)
  - administration procedure (e.g., individual vs. group)
  - time constraints (speed vs. power)
  - intended purpose (criterion vs. norm referenced)
Psychological Tests: Criterion Referenced

- Criterion referenced tests are used to evaluate an individual’s test score with reference to a set standard

- Examples of such tests include:
  - to earn a black belt in karate you must demonstrate a black-belt level of proficiency in karate
  - to drive a car you must take a driving test and perform to the satisfaction of the examiner
  - to conduct research with human participants at a university you must pass an ethics-oriented questionnaire

- Criterion referenced tests are typically used to gauge achievement or mastery, so they are sometimes called *mastery tasks*
Psychological Tests: Norm Referenced

- Norm referenced tests are used to compare a person’s test score with scores from a reference sample.
- The reference sample is considered to be representative of the population of interest.
- A person’s test scores is compared against the average score that would be obtained if the test were given to all members of the population of interest.
- Intelligence tests are norm referenced tests.
- In the US, college entrance exams like the Scholastic Achievement Test (SAT) are also norm referenced tests.
The main purpose of psychological measurement is to measure individual differences.

Psychological tests must be capable of:

1. Comparing the behaviour of different people—known as *interindividual differences*.
2. Comparing the behaviour of the same people at different times in different contexts—known as *intra-individual differences*.

The goal of psychological measurement is to quantify interindividual and/or intra-individual differences.

These individual differences contribute to test score variability—the currency of psychometric analysis.
Individual Differences

- The study of individual differences is known as *differential psychology*

- Historically, there has been a perception that psychometrics is the province of differential psychology

- This is because Francis Galton—the father of psychometrics—was a differential psychologist

- However, this is a gross misconception

- All psychologists measure behaviour and psychological attributes

- Therefore *all* psychologists must have an understanding of psychometrics

Francis Galton (1822–1911)
Challenges in Measurement

- There are many factors that can negatively affect the interpretation of test scores as valid.
- The textbook briefly reviews some of these factors in the designated chapter for this lecture e.g.,
  - demand characteristics
  - social desirability effects
  - malingering
  - experimenter bias
- I will cover this topic in more detail in subsequent lectures.