

Research Methods

LECTURE 3 VALIDITY IN PSYCHOLOGICAL RESEARCH

Measures

- ▶ Choosing appropriate measures for your study is another vital aspect of study design
- ▶ Need to think carefully about what it is you **actually** want to measure (think about your operational definition!) and what the best way of measuring this is...
 - E.g.,
 - ▶ Stress
 - ▶ Memory
 - ▶ Intelligence

Measures

Measures capturing psychological constructs are expected to be **reliable and valid**



Examples of Psychological Measures

► Psychometric tests

- IQ test
- Personality tests
 - Five Factor Personality Test
 - Openness to experience, Conscientiousness, Extraversion, Agreeableness, Neuroticism

Psycho = Mind
Metric = Measure

► Questionnaire measuring psychological distress

► Questionnaire measuring attachment style

Psychometric Standards

► Key measurement concepts:

1) **RELIABILITY:**

- Is the measure consistent and stable over time?

2) **VALIDITY:**

- Is it an accurate and precise measure?
- i.e., does it measure what it claims to measure?

► **STANDARDISATION**

e.g., exploring fear of stats in students...



Which would be the most reliable/valid way?

How could you explore this issue?

Measures

- ▶ Choosing an **existing** reliable and valid scale is always the best option:
 - ▶ You know that you are measuring something in a robust way (you haven't just made it up!)
 - ▶ You have existing data from other studies to compare your findings with
 - ▶ Sensitive to change within studies (if appropriate)
 - ▶ Allows for stronger statistical analyses

Impact of Event Scale (Horowitz et al., 1979)

(Tick one box from each row)	Not at all	Rarely	Sometimes	Often
I thought about it when I didn't mean to	☐	☐	☐	☐
I avoided letting myself get upset when I thought about it or was reminded of it	☐	☐	☐	☐
I tried to remove it from my memory	☐	☐	☐	☐
I had trouble falling asleep or staying asleep because of pictures or thoughts about it that came into my mind	☐	☐	☐	☐
I had waves of strong feelings about it	☐	☐	☐	☐
I had dreams about it	☐	☐	☐	☐
I stayed away from reminders of it	☐	☐	☐	☐
I felt as if it hadn't happened or wasn't real	☐	☐	☐	☐
I tried not to talk about it	☐	☐	☐	☐
Pictures about it popped into my mind	☐	☐	☐	☐
Other things kept making me think about it	☐	☐	☐	☐
I was aware that I still had a lot of feelings about it, but I didn't deal with them	☐	☐	☐	☐
I tried not to think about it	☐	☐	☐	☐
Any reminder brought back feelings about it	☐	☐	☐	☐
My feelings about it were kind of numb	☐	☐	☐	☐

RELIABILITY

CONSISTENCY
OR STABILITY OF
A MEASURE

Test-retest reliability

CONSISTENCY OF MEASUREMENT OVER TIME

- ▶ Assessed by administering the measure to the same group of participants on two or more separate occasions and correlating the sets of scores
- ▶ If the construct being measured is stable, scores should be consistent over time
- ▶ E.g., you would not expect the results of a test of IQ or personality traits to be different if measured more than once

Examples of The Big Five Inventory Personality measure

In General, I am someone who...	Disagree Strongly 1	Disagree a little 2	Neither agree nor disagree 3	Agree a little 4	Agree strongly 5
Is talkative					
Tends to find fault with others					
Does a thorough job					
Is depressed, blue					
Is original, comes up with new ideas					
Is reserved					
Is helpful and unselfish with others					
Can be somewhat careless					
Is relaxed, handles stress well					
Is curious about many different things					
Is full of energy					
Starts quarrels with others					
Is a reliable worker					

Internal Consistency

CONSISTENCY OF MEASUREMENT WITHIN THE TEST ITSELF

- ▶ All items on a test should measure the same thing
- ▶ Items should correlate highly with each other

Measure of the homogeneity of the test - i.e., the degree to which the various parts of an instrument measure the same variable

- ▶ Cronbach's alpha – response to each item is correlated with the response to every other item

Example: A scale of Extraversion

Very much so Moderately so Somewhat so Not at all so

I feel at ease with people.
 I start conversations.
 I take charge.
 I am the life of the party.
 I am skilled in handling social situations.
 I make friends easily.
 I talk to a lot of different people at parties.
 I know how to captivate people.
 I do not mind being the centre of attention.
 I feel comfortable around other people.

Response to each item should be correlated with the response to every other item – otherwise not tapping into the same thing

Reliability

Interpretation of Cronbach's Alpha

Certain benchmark values have been proposed in order to interpret Cronbach's alpha:

$\alpha < .60$ = poor!
 $.60 < \alpha < .70$ = adequate
 $.70 < \alpha < .80$ = good
 $.80 < \alpha < .90$ = very good!
 $\alpha > .90$ = perhaps too good...

Other forms of reliability...

► SPLIT-HALF RELIABILITY

- Items divided randomly
- Internal consistency between two sets of scores

► INTER-RATER RELIABILITY

- Extent to which two different raters agree in their evaluations
- Consistency of measurement when different people observe the same event or score the same test

VALIDITY

HOW WELL A TEST
ACTUALLY
MEASURES WHAT IT
IS DESIGNED TO
MEASURE

Face Validity

The test looks plausible at face value

- ▶ i.e., does it **look like** it is measuring what it claims to be measuring?
- ▶ E.g., Life Orientation Test (Scheier & Carver, 1985): a measure of dispositional optimism
 - ▶ In uncertain times I usually expect the best
 - ▶ If something can go wrong for me it will
 - ▶ I'm always optimistic about my future
 - ▶ I hardly ever expect things to go my way

Construct Validity

Does the test successfully measure the psychological construct it is designed to measure?

- ▶ E.g., intelligence
- ▶ Assessed by looking at the relationship between test scores and other behaviours that it should be related to
 - ▶ Do scores on a new IQ test relate to academic performance?
 - ▶ Does a measure of psychological distress report similar scores to a psychiatric interview?

Construct Validity

- ▶ Should be related to measures of the same construct – **convergent validity**
- ▶ Should be unrelated to measures of unrelated constructs – **discriminant validity**

Content Validity

- ▶ Whether the items on a test measure all the knowledge or skills that are assumed to underlie the construct of interest
- ▶ E.g., does a measure of intelligence capture all types of intelligence or just certain types?
 - ▶ E.g., problem-solving ability; numerical reasoning?

Criterion-related Validity

The ability of test scores to correlate with meaningful criterion measures

- ▶ Are they able to predict future outcomes?
- ▶ The test should predict present or future outcomes that are influenced by the construct being measured!
- ▶ E.g., does a measure of IQ successfully predict future performance/success in that job?
- ▶ E.g., do scores on a measure of conscientiousness predict future academic performance?

Measures and data.....

Recap: Measures

- ▶ Choosing an **existing** reliable and valid scale is always the best option:
 - ▶ You know that you are measuring something in a robust way (you haven't just made it up!)
 - ▶ You have existing data from other studies to compare your findings with
 - ▶ Sensitive to change within studies (if appropriate)
 - ▶ Allows for stronger statistical analyses

BUT BEWARE!: Scales and sub-scales

- ▶ Some psychological measures may include sub-scales
- ▶ Sometimes we can either add up all the items for a total score, or create separate sub-scales for each
- ▶ Sometimes we have to keep each sub-scale **separate** from the others

e.g., Impact of Event Scale (Horowitz et al., 1979)

- ▶ **INTRUSION SUB-SCALE:** Seven items measure intrusive thoughts – “Intrusion” about a stressful event (0-35)
“I thought about it when I didn't mean to...”
- ▶ **AVOIDANCE SUB-SCALE:** Eight items measure attempts to avoid those thought – “Avoidance” (0-40)
“I stayed away from reminders of it”
- ▶ Can be combined to create a “Global Distress Scale” (0-75)

(Tick one box from each row)	Not at all	Rarely	Sometimes	Often
I thought about it when I didn't mean to	☐	☐	☐	☐
I avoided letting myself get upset when I thought about it or was reminded of it	☐	☐	☐	☐
I tried to remove it from my memory	☐	☐	☐	☐
I had trouble falling asleep or staying asleep, because of pictures or thoughts about it that came into my mind	☐	☐	☐	☐
I had waves of strong feelings about it	☐	☐	☐	☐
I had dreams about it	☐	☐	☐	☐
I stayed away from reminders of it	☐	☐	☐	☐
I felt as if it hadn't happened or wasn't real	☐	☐	☐	☐
I tried not to talk about it	☐	☐	☐	☐
Pictures about it popped into my mind	☐	☐	☐	☐
Other things kept making me think about it	☐	☐	☐	☐
I was aware that I still had a lot of feelings about it, but I didn't deal with them	☐	☐	☐	☐
I tried not to think about it	☐	☐	☐	☐
Any reminder brought back feelings about it	☐	☐	☐	☐
My feelings about it were kind of numb	☐	☐	☐	☐

BUT BEWARE (AGAIN...)

- ▶ Some scales contain items which are negatively worded and need to be **reverse-coded** prior to analysis (to prevent demand characteristics/lazy responses)
 - ▶ E.g., Life Orientation Test (Scheier & Carver, 1985): a measure of dispositional optimism
 - ▶ In uncertain times I usually expect the best
 - ▶ If something can go wrong for me it will
 - ▶ I'm always optimistic about my future
 - ▶ I hardly ever expect things to go my way

Negatively worded items

Types of data

- ▶ Consideration of the type of data is important in study design
- ▶ Can often measure your operational definition through one of two types
 - ▶ Categorical
 - ▶ Continuous
- ▶ So... How do you decide??

Types of data

- ▶ Two broad types of data:

▶ Categorical

- ▶ E.g., gender, personality type
- ▶ Limited number of groups to which people can belong/can only belong to one group

▶ Continuous

- ▶ E.g., age, scores on a stress scale, seconds taken to complete a task
- ▶ Can take a vast range of values often from zero to an infinite number of different values

You can turn continuous data into categorical but not vice versa!

Why does it matter?

- ▶ Will influence
 - ▶ The quality of your results,
 - ▶ The type of conclusions you are able to draw from your results
 - ▶ The type of statistical analysis you are able to conduct (as you will learn in the next few weeks and beyond...)

Brief into to data

► Categorical data

- The results showed that a higher proportion of people in Group B (60%) were anxious compared to those in Group A (10%)

► Continuous data

- Can use mean scores
- The results showed that the mean anxiety score of participants in Group A was 2.3 and in Group B was 24.7 (measured on a scale of 0-30)
- Enables more sophisticated option for statistical analyses

Scales of measurement

Scale type	Properties of Scale
NOMINAL	Records differences A is red, B is blue A is extravert, B is introvert
ORDINAL	Records differences and magnitudes A is larger than B 1 st , 2 nd , 3 rd place (but not how many seconds between each place)
INTERVAL	Records differences and magnitudes and has equal intervals A is two units larger than B Temperature – equal units of measurement
RATIO	Records differences and magnitudes, has equal intervals and a true zero A is twice as large as B Race finish times – timed from a real zero mark

Summary

- Choosing appropriate measures for your study is another vital aspect of study design
- Choosing an **existing** reliable and valid scale is always the best option
- Scales should aim to be reliable and valid
- Beware scales and sub-scales
- And beware negatively worded items!
- Choose your scale and, subsequently, type of data carefully...



Questions?


