

MSc Applied Research
Ethics and Analysis (AREA)

Lecture 2: Questionnaire and Scale Data

Dr Kate Williams

Introduction

- To assess human characteristics, psychologists turn to certain measuring instruments:
 - Scales
 - Psychological tests
 - Questionnaires
- They tally up scores or examine responses.
- Choosing appropriate measures for your study is a vital aspect of study design.

Psychological Measures

- Method for measuring individual differences related to some psychological concept, based on a sample of relevant behaviour in a scientifically designed and controlled situation.
- E.g.,
 - Psychometric tests (E.g., IQ test)
 - Personality tests
 - A questionnaire measuring psychological distress
 - A questionnaire measuring attachment style

Psychological Scales and Tests

- Scales and tests aimed at measuring characteristics such as: attitudes, ability, intelligence, or personality are thought of as permanent measures and are seen as technical tools.
- They take a long time to develop and require expert knowledge to devise.

Scale Items

- E.g., in an attitude or personality scale, items in the scale often consist of a statement
 - “I am often late to lectures”
- The respondent would then answer this statement on a response scale:

Strongly
Agree

Agree

Neither
Agree or
Disagree

Disagree

Strongly
Disagree

This is known as a Likert Type scale (Likert, 1932)

Scale Items

- Each item measures a small aspect of a persons characteristic.
- A persons overall score from the scale is the sum of all responses on the scale:
 - Their overall score then depends on how much they tend to lean towards one end of the scale or the other.
 - E.g., Extroversion – Introversion

Psychological Scales and Tests

- Could be used for:
 - Selecting employees
 - Measuring levels of depression
 - Measuring a Cognitive variable

Responses may be used to examine differences between groups of people or examine differences within same group at different time points.

Psychological Scales and Tests

- Tend to use closed questions
 - E.g., Do you attend your lectures?
 - Always, Often, Sometimes, Never
- This allows data to be analysed *quantitatively*.
- Disadvantages of closed questions:
 - Constraint of responses
 - Misinterpretation
 - Unrealistic
 - Lacks rich information
 - *Why* did they give that response

Issues to consider

- Need to be careful that response choices do not overlap – if only one answer is expected then must ensure other answers cannot be circled.

Social Desirability effects

- Some people may feel pressure to respond a certain way even if it does not truly represent how they feel
- In an attempt to gain genuine responses, some measures can be cleverly disguised: if the responder is not aware of true intentions of the measure their answers are likely to be honest.

Social Desirability effects

- Disguising the measure is likely to be effective when responses involve disclosing highly sensitive, embarrassing, or potentially damaging information.
- *Can anyone see any issues with this?*
- *If so, what could you do about these issues?*

Types of Scales

- Equal appearing intervals (Thurstone, 1931)
 - Tick agreement from set of statement options.
 - Then given score equivalent to the strength of the statement if they agree with that statement (known as the Scale Value).
 - Involves: producing large number of statements, asking panel of judges to rate each statement, take mean value of all judges ratings, reject statements where judges' individual values have a high variance.

Types of Scales

- Summated ratings (Likert, 1932)
 - Set of statements which are responded to using a fixed response scale of choices.
 - ½ the items are favourable & ½ unfavourable
 - To avoid response set
 - Need to reverse some scores
 - Score for each item is summed to give overall score.
 - Can contain items not obviously measuring the attitude directly (Known as *Diagnostic*).

Types of Scales

- The Semantic differential scale (Osgood, Suci, & Tannenbaum, 1957).
 - P selects a point between opposite adjectives about the position they feel on the scale.

Exercising for 30 minutes every day would be....

Beneficial _ _ _ _ _ _ _ Harmful

Enjoyable _ _ _ _ _ _ _ Boring

Types of Scales

- The Visual Analogue Scale (VIS)
 - Effective for measuring characteristics that do not neatly fit in to one of two categories
 - E.g., Mood – you may not be high or low but somewhere in-between.

Indicate on the line your mood level right now

Low | - - - - - | High

Psychological Scales: Issues

- **Response Bias:** Harder to *disagree* with statements compared to agreeing.
- To deal with this confound, items in a scale need to be a mixture of positive and negative statements
 - Thus, the respondent needs to think carefully before answering.

Psychological Scales: Issues

- **Interpretations:** Some responders may think that you as the research believe in the statements on the scale (tricky if the items are quite sensitive/extreme)
 - Thus, it's a good idea to give clear explanations at the start: this is also another reason for including positive & negative items in a scale.

Psychological Scales: Issues

- **Demand Characteristics:** Participants may try to interpret the aim of the questions in an attempt to work out what is expected of them and attempt to answer accordingly.
- **Social Desirability:** Attempting to guess what is regarded as a socially acceptable answer and responding accordingly in order to be seen as 'good' or a 'good person', or in an attempt to please the researcher.

Scale construction: What to avoid

- Phrasing of items need to be clear and unambiguous, so the person answering will have no doubt what answer to give.
- Ideally want all responders to interpret the scale items in the same way.
 - In reality we know this does not always happen.
- What we can do to alleviate this is ensure certain things are avoided in order to maximise clarity.....

Scale construction: What to avoid

1. Complex Items
2. Technical Terms
3. Ambiguity
4. Double-barrelled questions
5. Double Negatives
6. Emotive Language
7. Leading Questions
8. Invasion of Privacy
9. Unbalancing
10. Sensitivity

What features should a good Psychological Measurement Scale include?

- Wide *Discriminability*
- Highly *Reliable*
- Be supported by tests of *Validity*
- Should be *Standardised*

Reliability

- Consistency of a measure: across different testing's (external) or within itself (internal).
- External Reliability: Stability across time.
 - Consistently producing the same outcome even on different occasions.
 - Measured with *Test-Retest Reliability*
 - Testing the **same** group of people twice (or more).
 - If the two sets of scores correlate, the tests is deemed as having high external reliability.

Reliability

- Internal Reliability: Consistency within the test. Answers to similar items in the test should be responded to in a similar way.
 - E.g., if a person has a high total neuroticism score then they should score in the neuroticism direction on each of the items in the scale.
 - Response to each item should be correlated with the response to every other item – otherwise not tapping into the same thing!

Checks for Internal Reliability

- Split-half method: The several items of the scale are divided randomly into 2 halves.
- If the test is reliable then scores for each half should be similar.
- Correlations are used to assess similarity
- However – how do you decide which items go in which half?

Checks for Internal Reliability

- Cronbach's alpha: Also called the coefficient's alpha is more widely used.
- It is related to the mean correlation between each pair of items and the number of items in the scale.
- Low alpha values indicate an unreliable test and emerge when Ps vary a lot on individual items relative to their overall variance.

Cronbach's alpha

- Good reliability of a scale is represented with alpha values between .75 and 1.
- As a general rule, any scale should have a Cronbach's alpha value of at least .7.
- Drawback: Even if a scale has a high alpha, individual items may still be poorly correlated with the others.
 - What can we do.....?

Item analysis

- To increase the reliability of a scale we can remove items causing the low reliability.
- When running a reliability analysis in SPSS, a table in the output called *Item-Total Statistics* is displayed.
 - In this table there is a column called 'Cronbach's alpha if Item deleted'.
 - We select the item that would most increase the alpha and simply delete it from the scale.

Item analysis

- We would continue to do this until a satisfactory alpha was achieved.
- Drawbacks: Ideally there should still be at least 10 items left in our scale after we have finished the deleting process – what if we need to delete more?
- Also, we cannot just delete items without paying attention to their content – we may end up altering the underlying construct.

Validity

- Does it measure what it intends to measure?

Face validity: test is plausible at face value and clear to everyone what it measures: the test 'makes sense'.

Content validity: Is it representative of the area it intends to cover – experts needed.

- Whether items on a test measure all the knowledge or skills that are assumed to underlie the construct of interest

Criterion validity: ability of test scores to correlate with meaningful criterion measures.

- Test should predict present or future outcomes that are influenced by the construct being measured

Criterion Validity

- Two specific types:
 - Concurrent validity: When a new measure is validated by comparison with an existing measure of the construct.
 - Predictive validity: When a prediction is accurately made from the measure.

Validity

Construct Validity: does it successfully measure the psychological construct it is designed to measure?

- Whole scientific process of establishing a psychological construct actually exists or at least is theoretically sound that it fits surrounding theory.
- Not a one-off procedure that gives a numerical value – it is the development of evidence for a hypothetical psychological construct through rigorous testing.
- Should be related to measures of the same construct – *convergent validity* – and it should be unrelated to measures of unrelated constructs – *discriminant validity*

Standardisation

- Uses measurement norms:
- E.g., you score 109 on an IQ test but your friend scores 118. What can you conclude from this? Your friend did better than you on the IQ test.
 - But in general did you both do well?
 - Is the difference between your scores large or small in the grander scheme of things?
 - How did other people like you perform?

Standardisation

- A test norm would give you these answers.
- It gives you a fair way of comparing people.
- Your test should be administered to a large sample of the target population (with means & standard deviations established) in order to make norm comparisons.
- This gives the overall average & the extend to which people vary around the average.

Standardisation

- E.g., IQ scales are adjusted until the mean is 100 and the average variation (standard deviation) is 15 points.
- Given that Psychometric tests are used on an applied basis (e.g., job selection, education, etc.) they must not discriminate a particular group or individuals in any way.
- Standardisation is therefore important both ethically and scientifically.

Normal Distribution

- Psychometric tests are usually adjusted until they resemble normal distribution.
- This can be seen as a positive in that it allows us to conduct powerful statistics.
- However: not all characteristics are normally distributed – attitudes towards some issues, for example, may display polarised opinions.

Standardisation

- Furthermore, can tests standardised on one particular population apply to another?
- What about cross-cultural research?
 - A test that's been standardised for one population may not be applicable to another.

Measures

Choosing reliable and valid scale is the best option:

- Measuring something in a robust way (not made it up!)
- Have existing data from other studies to compare your findings
- Sensitive to change within studies (if appropriate)
- Allows for stronger statistical analyses

Be Aware: Sub-scales

- Some psychological measures may include sub-scales
- Sometimes can either add up all items for total score, or create separate sub-scales for each
- Sometimes have to keep each sub-scale separate from 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 thoughts “avoidance” (0-40)
“I stayed away from reminders of it”
- Combined to create a “Global distress Scale” (0-75)

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	☐	☐	☐	☐

Beware again: Reverse-coding

Some scales contain items which need to be reverse-coded (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**

Choosing a Measure

- 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....
- You will need to consider the type of data your study will produce.
- Can often measure your operational definition through data type.

Questionnaires

- Gather information
- Typically used in surveys
- Tend to be constructed for a specific research topic
- Tend to focus on current opinion or patterns of behaviour
- Makes use of various types of scales

Scales and Questionnaire Data in Intervention research: changing health behaviour:

- Many of the risk factors for disease and illness are behavioural e.g.,
 - eating behaviour
 - smoking
 - lack of exercise
- Ability to change behaviour key to tackling many of the health issues in society today.

What is an intervention?

- **Interventions** in psychology are planned actions aimed at bringing about change in people's attitudes, beliefs and/or behaviours
- A wide range of intervention strategies exist
- Often involves structured activities used to modify behaviour, emotional state, or feelings

Assessing interventions

- Interventions must be based on sound empirical evidence
- Clinicians must be able to justify choice of treatment and show that they are able to deliver it effectively
- Use of psychological measures to achieve this
 - Collect scale and questionnaire data before and then after intervention to assess any change.

Interventions in Health Psychology

Must follow stringent guidelines for assessing clinical effectiveness as any medical treatment

Medical Research Council (MRC) framework

Randomised controlled trial