

Basic Concepts

Concept of a Measure

Questions

Responses

Scales of Measurement

Scaling

Scale Development

Concept of a Measure

Psychological Attributes

(also underlying constructs, psychological concepts, latent variables)

Psychological states, attitudes, abilities, personality characteristics (individual differences), or values that usually are not directly observable

Examples: anxiety, openness, job satisfaction, intelligence, fairness

Concept of a Measure

Psychological Test

"systematic procedure for comparing the behavior of two or more people" (Cronbach, 1960; p. 21)

Concept of a Measure

Psychological measures attempt to quantify the underlying psychological attribute

Numeric values used to represent greater or lesser degree of the psychological attribute

These numeric values must be meaningful enough to add, multiply, and use in statistical analyses

Concept of a Measure

Operational Definition

The psychological measure which is directly observed by the researcher or practitioner

Example: Spielberger's State-Trait Anxiety Inventory

Questions

Decide what psychological attributes to measure – driven by research or application objective

Decide general type of measure (e.g., quantitative, open ended, self-report, observational, reaction time)

Re-use existing measure or create a new one?

Questions

Desirable question properties (whether evaluating existing measure or creating a new one)

Clear relevance to the construct (consult prior research and theory)

Effective wording

- Simple, understandable, familiar vocabulary
- Brief (< 20 words often recommended)
- Avoiding ambiguities and vagueness
- Avoid idiomatic expressions, slang, abbreviations
- Cultural universality and sensitivity
- Carefully consider ordering (interpretation context is established by first questions)

Questions

Particular Wording Hazards

Leading questions

“Do you get headaches **frequently** and, if so, how often?” vs. “Do you get headaches **occasionally** and, if so how often?” (Loftus, 1979)

Double negatives

“I **never forget** to exercise”

“Double-barreled” questions

“I favor more gun laws **or** severe punishments?”

Low relevance to or low frequency for the respondents

“I enjoy the thrill I get from gaming”

“How afraid are you of talking to other people?”

“Loaded” questions (e.g., presuming, emotionally or politically charged, provocative)

Questions

Be careful, slight variations in item wording can have a large impact in responses. In a 2003 Pew Research Center survey...

When people were asked whether they would “favor or oppose taking military action in Iraq to end Saddam Hussein’s rule,” 68% said they favored military action while 25% said they opposed military action. However, when asked whether they would “favor or oppose taking military action in Iraq to end Saddam Hussein’s rule even if it meant that U.S. forces might suffer thousands of casualties,” responses were dramatically different; only 43% said they favored military action, while 48% said they opposed it.

Source: Pew Research Center <http://www.pewresearch.org/methodology/u-s-survey-research/questionnaire-design/>

Scales of Measurement

How do you want respondents respond to the question?

Responses can take many possible forms: most often in psychology we think about trying to quantify some psychological construct or attribute

Measurement scaling is the idea of linking a psychological attribute to a quantity

Scales of Measurement

Stevens (1946) proposed four scales or “levels” of measurement

Nominal – values **cannot be rank ordered**, e.g. religious group

Ordinal – values **can be rank ordered** but **do not have** equal distance among values e.g., “short” “tall” “taller”

Interval – values **can be rank ordered**, **do have** equal distances between values, and there is **no absolute zero** point, e.g. Fahrenheit temperature

Ratio – values **can be rank ordered**, **do have** equal distances between values, and **there is an absolute zero** point, e.g. Kelvin temperature

Scales of Measurement

Stevens originally proposed the classifications to help decide what statistical analysis should be used

Generally not regarded as definitive for analysis choice, but useful for thinking about how attributes are quantified

Common practice in social sciences is to use

- normal theory tests (e.g., ANOVA, t-tests, correlation) when the dependent variable is **interval, ratio, and ordinal with about five or more values**
- categorical data analyses (e.g., chi-square) when the dependent variables is **nominal or binary ordinal variables**
- special ordinal tests (e.g., loglinear, ordinal regression) when the dependent variable has **3 or 4 ordinal values**

Questions

Reverse worded items

To avoid *acquiescence bias* in which respondents be more likely endorse the agree side of the scale, most measures include some questions worded in the opposite direction

e.g., (frequency rating) “I am always happy” vs. “I am never happy”

Common to many measures, often recommended, but some potential problems (more on this topic later in the course)

Scales of Measurement

Dichotomous vs. continuous

Common examples of dichotomous scale: yes/no on survey, correct/incorrect on an ability test, symptom/behavior observed vs. not observed (nominal)

Examples of continuous scale: 7-point scale of agreement, frequency of a reported behavior (e.g., drinks per week), attitude/personality/symptomatology scale total score (treated as ordinal, interval, ratio)

Scales of Measurement

History

Tradition of psychophysics starting in 1800s (Gustav Fechner, Wilhelm Wundt) attempted to link behaviors, reactions, perceptions to numerical measurements

Historical approaches to self-report response measurement

- Bogardus Social Distance Scale
- Thurstone Equally appealing interval scale
- Guttman scale
- Semantic Differential (Osgood, Suci, Tannenbaum, 1957)
- Graphic scale, a variant on graphic scale

Scaling

Intuitive-Interval Scales (Bogardus Social Distance Scale, 1925)

Indicate which relationships you would be willing to have with each of the groups listed. Do not give your reactions to the best or the worst members you have known, but think of the picture or stereotype that you have of the whole group. In each column make an "X" after each of the relationships to which *you would be willing* to admit a member of that group.

	English	Swedes	Poles	Mexican	Koreans
(1) To close kinship by marriage.					
(2) To my club as a personal friend.					
(3) To my street as a neighbor.					
(4) To employment in my occupation.					
(5) To citizenship in my country.					
(6) As a visitor only to my country.					
(7) Would exclude from my country.					

Figure 12. *Bogardus Social-Distance Scale Format*

Source: Gorden, R. L. (1977). *Unidimensional scaling of social variables: Concepts and procedures*. New York: Free Press.

Scaling

Thurstone's Equally-Appearing Interval Scale (1928)

Developed a more systematic determination of the ranked alternatives by using an elaborate process which judges sorted many cards with possible option wordings into piles

Source: Gorden, R. L. (1977). *Unidimensional scaling of social variables: Concepts and procedures*. New York: Free Press.

Scaling

Thurstone's Equally-Appearing Interval Scale (1928)

As described by Allport (1935)

1. Specify the attitude variable to be measured.
2. Collect a wide variety of opinions relating to it from newspapers, books, or individuals
3. Assemble on cards approximately one hundred such typical opinions.
4. Require at least 200 to 300 judges to sort these cards into piles (eleven being the convenient and commonly used number), each pile representing equidistant degrees of attitude according to each judge's estimation.
5. Calculate scale value for each of the items by computing the median of the scale values assigned to it by judges and the dispersion of the judgements around the median.
6. Retain such statements that have small dispersions and are on the whole equally spaced. Give approximately equal representation to each of the intervals secured. Clarity and brevity of wording my furnish additional bases for selection.
7. In applying the scale, the subject checks every statement with which he or she agrees; the score is the mean scale value for all the endorsed statements

Source: Gorden, R. L. (1977). *Unidimensional scaling of social variables: Concepts and procedures*. New York: Free Press.

Scaling

Guttman scaling (cumulative, scalogram; 1929; 1959)

“With which *one* of these statements concerning postwar relations with Russia do you come closest to agreeing?”

- (1) It is very important to keep on friendly terms with Russia, and we should make every possible effort to do so.
- (2) It is important for the U.S. to be on friendly terms with Russia, but not so important that we should make too many concessions to her.
- (3) If Russia wants to keep on friendly terms with us, we shouldn't discourage her, but there is no reason why we should make any special effort to be friendly.
- (4) We shall be better off if we have just as little as possible to do with Russia.

To test the assumption that the four statements actually indicated the degrees of favorableness intended, the researchers asked the question of a sample of 3,000 enlisted men in the following form:

“Do you agree with this statement?”

- (1) It is very important to keep on friendly terms with Russia, and we should make every possible effort to do so.
— Agree
— Disagree
— Undecided

Source: Gorden, R. L. (1977). *Unidimensional scaling of social variables: Concepts and procedures*. New York: Free Press.

Scaling

Guttman scaling (cumulative, scalogram; 1929; 1959)

Judges rate the items for relevance (yes, no)

1. Create a matrix of the yes/no results
2. Sort the items according to the judges who endorsed the most items as relevant
3. Then sort across the columns according to the items most endorsed from less to more
4. Select the subset of items most endorsed to use

	Item 1	Item 2	Item 3
Judge 1	Y	Y	Y
Judge 2	Y	Y	Y
Judge 3	Y	Y	N
Judge 4	Y	N	N

Scaling

Semantic Differential (Osgood, Suci, Tannenbaum, 1957)


Good . ____ . ____ . ____ . ____ . ____ . ____ . ____ . ____ .Bad

Strong . ____ . ____ . ____ . ____ . ____ . ____ . ____ . ____ .Weak

Fast . ____ . ____ . ____ . ____ . ____ . ____ . ____ . ____ .Slow

Scaling

A variant on the semantic differential is the graphic scale

Shy _____  _____ Outgoing

Scaling

Likert (1932) Type Scaling¹

I find psychometrics intimidating.

1	2	3	4	5
Strongly disagree	Disagree	Neither Agree nor	Agree	Strongly Agree

¹ Likert scaling is named after Rensis Likert, who pronounced his name as *lick' - uhrt*, although most people say *like' - uhrt*

Scaling

Likert-Type Scaling

There are several common variants of the Likert-type scale, such as frequency ratings

How often do you find psychometrics intimidating?

1	2	3	4	5
Never	Seldom	Sometimes	Often	Always

Scaling

Likert-Type Scaling

Likert scaling does a pretty good job of replicating an equally-spaced ordinal measure (better approximation of an interval scale) if numeric descriptors are carefully worded and symmetric

Better to use a more common (and familiar) set of descriptors than to try to develop something unique

Scaling

Likert-Type Scaling

A number of studies suggest the more response options up to as many as 11 improves reliability of the measure, although improvements tend to be greatest up to 7 to 9 options (e.g., Preston & Colman, 2000)

7-point Likert scales seem to be the most common in psychological research, followed closely by 5-point or 9-point scales

Scaling

Including neutral points vs. forced choice

Forced choice can be a “Yes” or “No” responses with no “Neither”

Or forced choice can occur with a multiple option scale, such as Likert-type

1	2	3	4
Strongly disagree	Disagree	Agree	Strongly agree

Scaling

Forced choice

Can lead to measurement errors or nonresponse errors if respondents do not truly agree or disagree

Can evoke feelings of resistance in respondents (Cronbach, 1960)

Resistance responses are reduced with more options (Highland & Berkshire, 1951)

If not balanced (e.g., more negative options), can be leading

More complex multidimensional forced choice approaches can be used to reduce response biases (Brown & Maydeu-Olivares, 2011)

Scaling

Don't Know's

That having a “don't know” option is often appealing to participants, they cause problems for data analysis, because, in nearly all instances, they have to be treated as missing data.

Many surveys include a “don't know” option, for self-completed questionnaires, they tend to allow more “don't know” responses,

But for interviews, interviewers can diplomatically suggest the respondent try to choose the option that comes closest to the respondent's opinion, and, usually successfully decrease “don't know” responses