LECTURE 17 SCALING TECHNIQUES

DR. GAURAV SANKALP

SCALING TECHNIQUES

LIKERT SCALE

A **Likert scale** is a psychometric scale commonly involved in research that employs questionnaires. It is the most widely used approach to scaling responses in survey research, such that the term is often used interchangeably with *rating scale*, or more accurately the **Likert-type scale**, even though the two are not synonymous. The scale is named after its inventor, psychologist Rensis Likert. Likert distinguished between a scale proper, which emerges from collective responses to a set of items (usually eight or more), and the format in which responses are scored along a range. Technically speaking, a Likert scale refers only to the former. The difference between these two concepts has to do with the distinction Likert made between the underlying phenomenon being investigated and the means of capturing variation that point to the underlying phenomenon. When responding to a Likert questionnaire item, respondents specify their level of agreement or disagreement on a symmetric agree disagree scale for a series of statements. Thus, the range captures the intensity of their feelings for a given item. A scale can be created as the simple sum questionnaire responses over the full range of the scale. In so doing, Likert scaling assumes that distances on each item are equal. Importantly, "*All items are assumed to be replications of each other or in other words items are considered to be parallel instruments*"

SAMPLE QUESTION PRESENTED USING A FIVE-POINT LIKERT ITEM

An important distinction must be made between a *Likert scale* and a *Likert item*. The Likert scale is the sum of responses on several Likert items. Because Likert items are often accompanied by a visual analog scale (e.g., a horizontal line, on which a subject indicates his or her response by circling or checking tick-marks), the items are sometimes called scales themselves. This is the source of much confusion; it is better, therefore, to reserve the term *Likert scale* to apply to the summed scale, and *Likert item* to refer to an individual item.

A Likert item is simply a statement which the respondent is asked to evaluate according to any kind of subjective or objective criteria; generally the level of agreement or disagreement is measured. It is considered symmetric or "balanced" because there are equal amounts of

positive and negative positions. Often five ordered response levels are used, although many psychometricians advocate using seven or nine levels; a recent empirical study found that a 5- or 7- point scale may produce slightly higher mean scores relative to the highest possible attainable score, compared to those produced from a 10-point scale, and this difference was statistically significant. In terms of the other data characteristics, there was very little difference among the scale formats in terms of variation about the mean, skewness or kurtosis.

The format of a typical five-level Likert item, for example, could be:

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

Likert scaling is a bipolar scaling method, measuring either positive or negative response to a statement. Sometimes an even-point scale is used, where the middle option of "Neither agree nor disagree" is not available. This is sometimes called a "forced choice" method, since the neutral option is removed. The neutral option can be seen as an easy option to take when a respondent is unsure, and so whether it is a true neutral option is questionable. A 1987 study found negligible differences between the use of "undecided" and "neutral" as the middle option in a 5-point Likert scale.

Likert scales may be subject to distortion from several causes. Respondents may avoid using extreme response categories (*central tendency bias*); agree with statements as presented (*acquiescence bias*); or try to portray themselves or their organization in a more favorable light (*social desirability bias*). Designing a scale with balanced keying (an equal number of positive and negative statements) can obviate the problem of acquiescence bias, since acquiescence on positively keyed items will balance acquiescence on negatively keyed items, but central tendency and social desirability are somewhat more problematic.

MULTIDIMENSIONAL SCALING

Multidimensional scaling (**MDS**) is a means of visualizing the level of similarly of individual cases of a dataset. It refers to a set of related ordination techniques used in

information visualization, in particular to display the information contained in a distance matrix. An MDS algorithm aims to place each object in *N*-dimensional space such that the between-object distances are preserved as well as possible. Each object is then assigned coordinates in each of the *N* dimensions. Unlike principal component analysis wherein most of the variance in the data is captured in the first axis with each subsequent axis containing progressively less information, axes in MDS are arbitrary and distance units along each axis do not reflect equal quantitative distances at other sections of the same axis. The number of dimensions of an MDS plot *N* can exceed 2 and are specified a priori. Choosing *N*=2 optimizes the object locations for a two dimensional scatterplot.

TYPES OF MULTIDIMENSIONAL SCALING

1. **Classical multidimensional scaling:** Also known as Principal Coordinates Analysis, Torgerson Scaling or Torgerson–Gower scaling. Takes an input matrix giving dissimilarities between pairs of items and outputs a coordinate matrix whose configuration minimizes a loss function called *strain*.
2. **Metric multidimensional scaling:** A superset of classical MDS that generalizes the optimization procedure to a variety of loss functions and input matrices of known distances with weights and so on. A useful loss function in this context is called *stress*, which is often minimized using a procedure called stress majorization.
3. **Non-metric multidimensional scaling:** In contrast to metric MDS, non-metric MDS finds both a non-parametric monotonic relationship between the dissimilarities in the item-item matrix and the Euclidean distances between items, and the location of each item in the low-dimensional space. The relationship is typically found using isotonic regression. Louis Guttman's smallest space analysis (SSA) is an example of a non-metric MDS procedure.
4. **Generalized multidimensional scaling:** An extension of metric multidimensional scaling, in which the target space is an arbitrary smooth non-Euclidean space. In cases where the dissimilarities are distances on a surface and the target space is another surface, GMDS allows finding the minimum-distortion embedding of one surface into another.

SELECTION OF APPROPRIATE ATTITUDE MEASUREMENT SCALE

Now that we have looked at a number of attitude measurement scales, a natural question arises: "Which is most appropriate?" As in the selection of a basic research design, there is no single best answer for all research projects. The answer to this question is relative, and the choice of scale will depend on the nature of the attitudinal object to be measured, the manager's problem definition, and the backward and forward linkages to choices already made (for example, telephone survey versus mail survey). However, several questions will help focus the choice of a measurement scale:

- Is a ranking, sorting, rating, or choice technique best?
- Should a monadic or a comparative scale be used?
- What type of category labels, if any, will be used for the rating scale?
- How many scale categories or response positions are needed to accurately measure an attitude?
- Should a balanced or unbalanced rating scale be chosen?
- Should a scale that forces a choice among predetermined options be used?
- Should a single measure or an index measure be used?

RANKING, SORTING, RATING, OR CHOICE TECHNIQUE

The decision whether to use ranking, sorting, rating, or a choice technique is determined largely by the problem definition and especially by the type of statistical analysis desired. For example, ranking provides only ordinal data, limiting the statistical techniques that may be used.

Monadic or Comparative Scale

If the scale to be used is not a ratio scale, the researcher must decide whether to include a standard of comparison in the verbal portion of the scale. Consider the following rating scale:

Now that you've had your automobile for about one year, please tell us how satisfied you are with its engine power and pickup.

Completely Dissatisfied	Dissatisfied	Somewhat	Satisfied	Completely Satisfied

[ ]	[ ]	[ ]	[ ]	[ ]

This is a monadic rating scale, because it asks about a single concept (the brand of automobile the individual actually purchased) in isolation. The respondent is not given a specific frame of reference. A comparative rating scale asks a respondent to rate a concept, such as a specific amount of responsibility or authority, in comparison with a benchmark—perhaps another similar concept—explicitly used as a frame of reference. In many cases, the comparative rating scale presents an ideal situation as a reference point for comparison with the actual situation. For example:

Please indicate how the amount of authority in your present position compares with the amount of authority that would be ideal for this position:

Too Much [ ] About Right [ ] Too Little [ ] What Type of Category Labels, If Any?

We have discussed verbal labels, numerical labels, and unlisted choices. Many rating scales have verbal labels for response categories because researchers believe they help respondents better understand the response positions. The maturity and educational levels of the respondents will influence this decision. The semantic differential, with unlabeled response categories between two bipolar adjectives, and the numerical scale, with numbers to indicate scale positions, often are selected because the researcher wishes to assume interval-scale data.

How Many Scale Categories or Response Positions?

Should a category scale have four, five, or seven response positions or categories? Or should the researcher use a graphic scale with an infinite number of positions? The original developmental research on the semantic differential indicated that five to eight points is optimal. However, the researcher must determine the number of meaningful positions that is best for the specific project. This issue of identifying how many meaningful distinctions respondents can practically make is basically a matter of sensitivity, but at the operational rather than the conceptual level.

BALANCED OR UNBALANCED RATING SCALE

The fixed-alternative format may be balanced or unbalanced. For example, the following question, which asks about parent-child decisions relating to television program watching, is a **balanced rating scale**:

Who decides which television programs your children watch?

- Child decides all of the time. [ ]
- Child decides most of the time. [ ]
- Child and parent decide together. [ ]
- Parent decides most of the time. [ ]
- Parent decides all of the time. [ ]

This scale is balanced because a neutral point, or point of indifference, is at the center of the scale.

Unbalanced rating scales may be used when responses are expected to be distributed at one end of the scale. Unbalanced scales, such as the following one, may eliminate this type of "end piling":

Completely Dissatisfied	Dissatisfied	Somewhat	Satisfied	Completely Satisfied
[ ]	[ ]	[ ]	[ ]	[ ]

Notice that there are three "satisfied" responses and only two "dissatisfied" responses above. The choice of a balanced or unbalanced scale generally depends on the nature of the concept or the researcher's knowledge about attitudes toward the stimulus to be measured.

USE A SCALE THAT FORCES A CHOICE AMONG PREDETERMINED OPTIONS

In many situations, a respondent has not formed an attitude toward the concept being studied and simply cannot provide an answer. If a forced-choice rating scale compels the respondent to answer, the response is merely a function of the question. If answers are not forced, the midpoint of the scale may be used by the respondent to indicate unawareness as well as indifference. If many respondents in the sample are expected to be unaware of the attitudinal

object under investigation, this problem may be eliminated by using a non-forced-choice scale that provides a "no opinion" category, as in the following example:

- How does the Bank of Commerce compare with the First National Bank? [ ]
- Bank of Commerce is better than First National Bank [ ]
- Bank of Commerce is about the same as First National Bank [ ]
- Bank of Commerce is worse than First National Bank [ ]
- Can't say [ ]

Asking this type of question allows the investigator to separate respondents who cannot make an honest comparison from respondents who have had experience with both banks. The argument for forced choice is that people really do have attitudes, even if they are unfamiliar with the banks, and should be required to answer the question. Still, the use of forced-choice questions is associated with higher incidences of "no answer." Internet surveys make forced-choice questions easy to implement because the delivery can be set up so that a respondent cannot go to the next question until the previous question is answered. Realize, however, if a respondent truly has no opinion, and the no opinion option is not included, he or she may simply quit responding to the questionnaire.

SINGLE MEASURE OR AN INDEX MEASURE

Whether to use a single measure or an index measure depends on the complexity of the issue to be investigated, the number of dimensions the issue contains, and whether individual attributes of the stimulus are part of a holistic attitude or are seen as separate items. Very simple concepts that do not vary from context to context can be measured by single items. However, most psychological concepts are more complex and require multiple-item measurement. Additionally, multiple-item measures are easier to test for construct validity. The researcher's conceptual definition will be helpful in making this choice.

The researcher has many scaling options. Generally, the choice is influenced by plans for the later stages of the research project. Again, problem definition becomes a determining factor influencing the research design.