

12

Developing Scales

The Likert Scale
The Thurstone Scale
The Q-Sort Scale
The Semantic Differential

Our emphasis from the beginning of the text has been on the **objective measurement of research variables**. Sometimes we are most interested in studying **subjective variables**: attitudes, feelings, personal opinions, or word usage. **How can we measure subjective variables objectively?** The answer is an instrument called a scale.¹

Dr. Martha Bergen used an adaptation of an existing scale² to measure the attitude of seminary professors toward using computers in seminary education.

Respondants [110 seminary professors serving at Southwestern Baptist Theological Seminary in 1988] were asked to read each question and decide to what extent they agreed or disagreed with each question. They were instructed to circle the appropriate number after each of the items. The rating scale was set up in a logical pattern using the numbers "1," "2," "3," and "4" to correspond with "strongly disagree," "disagree," "agree," and "strongly agree," respectively. Responses [from the 53 items] were totaled and evaluated to reveal which attitude/s was/were most prominent. . . .

A validation panel consisting of five experts in the areas of education, religious education, and computers was asked to rate the relevance and clarity of each question. Proper revisions and modifications were made as deemed necessary from the panel's critique and evaluation.

For the purpose of establishing reliability, a stratified random sample of ten seminary professors -- representative of the intended population -- was selected to respond to the questionnaire. The method of split-half correlation was used to determine the coefficient of internal consistency. . . .³

The result of the modifications was an **instrument which measured the strength of support (an attitude) of seminary professors for the use of computers** in the seminary education in 1988. The internal consistency coefficient, after applying the Spearman-

¹See Babbie, "Chapter 15: Indexes, Scales and Typologies," pp. 366-389; Nunnally, "Chapter 15: Attitudes and Interests," pp. 441-467; and Payne, "Chapter 8: The Development of Self-Report Affective Items and Inventories," pp. 164-200. **An excellent paperback dealing with this subject is Daniel J. Mueller, *Measuring Social Attitudes: A Handbook for Researchers and Practitioners*, (New York: Teachers College Press, 1986).**

²Bergen describes her instrument as "an adaptation of a 1986 dissertation [instrument] from North Texas State Univeristy. See Mitchell Drake Weir, 'Attitudes and Perceptions of Community College Educators toward the Implementation of Computers for Administrative and Instructional Purposes' (Ph.D. dissertation, North Texas State University, 1987), pp. 129-35. In May 1988 North Texas State University became the University of North Texas," 48

³*Ibid.*, 48-49. See also 57-62 for more detail.

Brown Prophecy Formula, was +0.75, a strong positive value (see Chapter 22).

A scale is an instrument which measures subjective variables. In this chapter we look at four major types of scales: the Likert (LIE-kurt), the Thurstone, the Q-sort and the Semantic Differential. Each of these important scale types provides the means to gather subjective data objectively.

The Likert Scale

The Likert scale is by far the most popular attitude scale type. A statement is followed by several levels of agreement: strongly agree, agree, no opinion, disagree, strongly disagree. This five-point scale is commonly used, but other scales, from four to ten points, can be used as well.²

Follow these steps to develop a Likert Scale for use in research. We will use the attitude “desire to learn” in college students to illustrate the process.

Define the attitude

The first step in designing an attitude scale is to define the attitude you want to measure. What does the attitude mean? What does “desire to learn” mean? If students do not have a desire to learn, what do they have? Perhaps, “desire to get a degree.” With these two end points we can begin to build a scale to differentiate between those who desire to learn, and those who merely want a credential.

In defining the attitude, we must choose which end of the scale will be positive, and which will be negative. The simplest way to do this is to assign the positive end of the scale to your attitude. For our example, we'll make “desire to learn” positive, and “desire to get a degree” negative.

Determine related areas

Having defined the end points of the scale, we next determine what attitudes, opinions, behaviors, or feelings might be related to each end of the scale. What kinds of things would reflect the positive side? The negative side? These related areas provide the raw material from which we'll develop attitudinal statements.

In what areas would “learn” and “degree” students differ? Here's my suggested list: doing homework, using the library, extra reading, free time discussion, meetings with professors, opinions concerning the meaning of a degree, and views on grades.

Write statements

Next, we will write statements that reflect positive and negative aspects of these areas. We've defined “positive” to mean “that which agrees with my position,” and “negative” means “that which disagrees with my position.” The statements, even though reflecting subjective variables, should be objective. That is, statements must not be systematically biased toward one position or the other. Students who really want merely to get a degree should have no trouble scoring low on the scale. They should tend to agree with statements reflecting “degree” and tend to disagree with statements reflecting “learning.” In the same way, students who really want to learn should tend to agree with “learning” statements, and tend to disagree with “degree”

²Mueller, “Chapter 2: Likert Attitude Scaling” and “Chapter 3: Likert Scale Construction: A Case Study,” 8-33.

statements.

Positive examples

Positive statements should be objective statements which are acceptable by those having the attitude, and just as unacceptable to those not having it. The following reflect these characteristics in regard to our attitude scale:

- I generally enjoy homework assignments and sometimes do more than the assignment requires.
- I frequently use library resources to go beyond the required reading.
- I believe a degree is empty unless it reflects my best efforts of scholarship.
- A late assignment, thoughtfully done, is more important than the loss in grade average.

Negative examples

Negative statements should be objective statements which are acceptable to those not having the attitude, and just as unacceptable to those having it. These statements coincide with the positive examples above.

- Homework assignments are designed to meet course requirements. It is impractical in time and energy to do more than is required.
- It is better to master the required reading than to dilute one's thinking with other authors.
- A degree is a credential for ministry and reflects, in itself, none of the extremes of scholarship some try to ascribe to it.
- It is better to turn in an assignment on time than to be docked for lateness to make it better.

Create an item pool

Continue writing items, both positive and negative, until you have an item pool at least **twice the size of your intended instrument**. If you plan to have 20 statements in your final scale, then create an item pool of 40 items.

Validating the items

Enlist a validation panel of 6-8 persons to evaluate each item. It is suggested that you have persons on the panel who represent both extremes of the scale. Have the panel rate each item on its clarity and potency in defining the attitude in question.

Rank

Rank order the evaluated items on **clarity and potency**. Choose an equal number

³Mueller states, "Five categories are fairly standard.... Some scale constructors use seven categories, and some prefer four or six response categories (with no middle category). All of these options seem to work satisfactorily. It should be noted in this regard that reducing the number of response categories reduces the spreading out of scores (reduces variance) and thus tends to reduce reliability. Increasing the number of response categories adds variance. As the number of categories is increased, a point is reached at which respondents can no longer reliably distinguish psychologically between adjacent categories [i.e., what's the difference between a 10 and an 11 on a 12-point scale? WRY]. Increasing the number of categories beyond this point simply adds random (error) variance to the score distribution" (pp. 12-13).

of positive and negative items from the best statements.

Formatting the Scale

Randomly order the selected statements. Use letters to indicate choices, such as “SD”, “D”, “A”, and “SA” rather than numbers. I recommend that you use four or six levels of response. Using an **even number** of responses forces respondents to mark the direction of their attitudinal tendencies – positive or negative. Mean scores for groups filling out the scale have more meaning in this less stable construction. Many Likert scales have 5 levels, with a “no opinion” center. This neutral middle option allows subjects an easy way to avoid considering the statement.

Write instructions

Write instructions which clearly explain how to select responses on the form. (See the finished example at the end of the chapter.) There are other ways to indicate the intensity of response. **Dr. Don Mattingly** (Ed.D., 1984) developed a scale for his dissertation which used the categories

Yes! Yes No No!

to indicate how strongly his subjects agreed or disagreed with statements concerning recreation ministry.

Scoring the Likert scale

The points given for each response depend on whether the statement is positive or negative. The person who “strongly agrees” with a “positive statement” gets the maximum points. One who “strongly disagrees” with a “positive statement” gets the minimum points. For a four-point scale, the scoring would be as follows for *positive statements*: SD=1, D=2, A=3, SA=4.

The person who “strongly agrees” with a negative statement gets the minimum number of points (1), while the one who “strongly disagrees” with a negative statement gets the maximum points (4). In our four-point example, the scoring for *negative statements* would be as follows: SD=4, D=3, A=2, and SA=1.

In this short 8-item example attitude scale (see end of chapter), subject attitude scores will range from a low of “8” ($8 \times 1 = 8$) to a high of “32” ($8 \times 4 = 32$). For a twenty-five item scale, this procedure yields scores ranging from 25 to 100. These scores can then be used to compare groups on the defined attitude.

The Thurstone Scale

The Likert Scale, which we just discussed, consists of statements that are all of **equal weight**. The subjects’ score results from adding together all of the scaled responses for all the statements. **The Thurstone attitude scale**, however, consists of statements which have a **range of weights from high (usually 11) to low (usually 1)**. Subjects select the attitudinal statements they agree with most. Their scores result from computing the average of the weights of the items selected.⁴ Use the following steps to develop weighted items for a Thurstone scale.

⁴See Mueller, “Chapter 4: Thurstone Scale Construction,” 34-46.

Attitude Toward Seminary Learning

INSTRUCTIONS: Read each statement below. Circle the letter which best describes your response to the statement. If you strongly disagree with the statement, circle SD. If you DISAGREE, circle D, AGREE, A, or STRONGLY AGREE, SA.

- | | |
|---|------------------------|
| 1. Homework assignments are designed to meet course requirements. It is impractical in time and energy to do more than is required. (-) | SD (D) A SA
= 3 pts |
| 2. A late assignment – thoughtfully done – is more important to me than the loss in grade average. (+) | SD D (A) SA
= 3 pts |
| 3. A degree is a credential for ministry and reflects, in itself, none of the extremes of scholarship some try to ascribe to it. (-) | (SD) D A SA
= 4 pts |
| 4. I generally enjoy homework assignments and sometimes do more than the assignment requires. (+) | SD D (A) SA
= 3 pts |
| 5. It is better to turn in an assignment on time, as it is, than to be docked for lateness to make it better. (-) | SD D (A) SA
= 2 pts |
| 6. I frequently use library resources to go beyond the required reading. (+) | SD (D) A SA
= 2 pts |
| 7. I believe a degree is empty unless it reflects my best efforts of scholarship. (+) | SD D A (SA)
= 4 pts |
| 8. It is better to master the required reading than to dilute one's thinking with other authors. (-) | SD D (A) SA
= 2 pts |

Red notations are not included on the form, but are included here to demonstrate the scoring of a completed form. This subject selects items as marked, which are scored according to statement type. This subject scored **23 points** on this scale (32 possible).

Very positive attitude!

Develop item pool

As in the development of Likert scale items, develop an item pool of attitudinal statements. Include statements that range from **extremely unfavorable** to **extremely favorable**, as well as neutral statements. An item pool of about 50 attitudinal statements is adequate.

Compute item weights

Compute a scale value (or “weight”) for each statement. This is done by having a panel of 10 or more judges rank each statement. This is done by having each judge read through all statements, and choosing the most positive. This statement is given 1 point. Most negative? (11 points). These statements are eliminated from the pool. Choose the two next most positive (2 points each). Two next most negative (10 points). Four next most positive (3 points). Four next most negative (9 points). After all judges have rank ordered the statements, **average weights** are computed by adding up all the points from all the judges for each item, and dividing by the number of judges. This average is the **item weight**. The item with the lowest weight is the most positive according to the panel of judges. The item with the highest weight is the most negative.

Rank the items by weight

Rank the items by item weight, **low (positive) to high (negative)**.

Choose Items by Equidistant Weights

Compose the final scale by selecting 20 to 25 statements whose **weights are approximately equidistant from each other throughout the entire scale**. If a 9-category favorableness scale was used by judges and if 22 items are to be selected for the final scale, the items will need to be picked at scale intervals of approximately .36. (There are eight units between 1.00 and 9.00; $8/22 = 0.36$). In fact, since no items will have median values as low as 1.00 or as high as 9.00, a slightly smaller interval size, perhaps around .33, should be used to select 22 equidistant items.⁵

If two items have the same weight, choose the item with the smaller standard deviation (see Chapter 16 for how to calculate standard deviation). In this way, the list of statements form a range of weights, as determined by the panel of judges.

Formatting the Scale

Place the selected statements in random order. **Do not include item weights on the instrument.**

Administering the Scale

Direct the subjects to read all statements in the instrument and **mark those with which they agree**. They may choose as many as they like. **See the example at the end of the chapter.**

Scoring

Compute the median (or mean) of the weights of the statements marked by the

⁵Mueller, p. 37

subject. This is the subject's score which reflects attitude on the theme.

Q-Methodology

It is difficult to rank order more than ten statements. But rank ordering attitudinal statements is a good way to gather subjective data on a given sample. The "Q-sort" is a procedure for rank ordering a large number of statements. Rankings of statements by two or more groups can then be compared.

One version of the Q-sort uses a physical set of boxes, numbered 1 through 11 (This is the same arrangement as that described for weighting Thurstone items). The procedure is usually applied when the number of statements to be ranked is greater than 40. The subject looks through a number of statements written on cards. Each card contains one statement.

The first time through, the subject selects the statement he agrees with the most. That item goes into box 1.

The subject then goes through the cards a second time and selects the statement he agrees with the least. This card is placed in box 11.

The next time through the cards, the subject selects two cards he agrees with the most, and places them into box 2.

Then he chooses the two cards he agrees with least in box 10. Then 4 cards in box 3 and 4 cards in box 9, and so forth, until he is left with the middle box (#6). All the remaining statements are placed in it. The researcher then assigns point values for each statement, 1-11, based upon the box into which they were placed.

After all subjects have placed the statements, averages are computed. Rank order statements for the group on the basis of their average values.

Semantic Differential

The semantic differential provides information on differences ("differential") in word usage ("semantics") in subjects. Osgood and Tannenbaum wrote the classic work on using the semantic differential, entitled *The Measurement of Meaning*.¹ The book is a detailed analysis of this powerful technique. We simply introduce the procedure here.

Osgood and Tannenbaum isolated three major dimensions of word meanings through the use of factor analysis. These dimensions are evaluative (good or bad), potency (strong or weak) and activity (fast or slow). Their book contains hundreds of adjective pairs relating to these three dimensions.

A subject is presented a sheet of paper with a single word or term at the top. Below this word are a number of adjectival pairs, separated by seven blanks. For example, the meanings associated with the term "my church" might be formatted like this:

My Church							
valuable	__	:	__	:	__	:	worthless
clean	__	:	__	:	__	:	dirty
bad	__	:	__	:	__	:	good
unfair	__	:	__	:	__	:	fair
large	__	:	__	:	__	:	small
strong	__	:	__	:	__	:	weak
deep	__	:	__	:	__	:	shallow
fast	__	:	__	:	__	:	slow
active	__	:	__	:	__	:	passive
hot	__	:	__	:	__	:	cold
	(1)	(2)	(3)	(4)	(5)	(6)	(7)

The first four adjective pairs measure the evaluative dimension; the next three measure potency; and the last three measure activity. *The numbers shown above are not*

¹Urbana: University of Illinois Press, 1957

printed on the instrument, but are shown here to help clarify the scoring procedure. Pairs which are reversed should be scored in reverse, so that positive is always (1) and negative (7) regardless of which side of the scale they appear.

Subjects check one blank between each pair indicating their opinion of the term on this scale. Blanks are scored 1-7, providing a numerical score for the meaning of the term in each dimension. Groups of subjects can then be compared on the three dimensions of meaning for any commonly used word. (Note: the numbering scale 1-7 is true only if the positive term is on the left; otherwise the scale is labelled 7-1). Results can be plotted in three dimensions — to provide a picture of semantic differences between two or more groups of subjects.

The Delphi Technique

The Delphi Technique, while having alternative forms and procedures, is essentially used to **determine consensus in a group of subjects**. The items around which this consensus is formed are **constructed from comments from the group itself**, thereby eliminating researcher bias in item creation.

Suppose a researcher is interested in defining the most important concerns of Sunday School teachers of youth in Tarrant Baptist Association churches. A letter would be sent to all youth teachers in Tarrant Association churches asking them to list their "major concerns" in teaching young people. Responses would most likely range from literature, to facilities, to youth attitudes, to parental problems, to . . . well, the list would be long. The "major concerns" from all responders would be analyzed for commonalities, and a list of key "major concerns" would be produced.

From this list of "major concerns" the researcher would create pairs of attitudinal statements, one positive and the other negative. For example, for the major concern of "youth literature," one might create the following pair of attitudinal statements.

- (+) The literature we use for teaching youth demonstrates an understanding of youth needs, and how the Bible addresses those needs.
- (-) The literature we use for teaching youth demonstrates a lack of understanding of youth needs, and provides little help in addressing those needs with the Bible.

Pairs of statements are created for each major concern.

Randomly select an equal number of **positive and negative statements** for inclusion in the Delphi instrument.

Construct an instrument in which statements are **randomly listed**.

Associate each with a **Likert type response**: Strongly Agree. . . Strongly Disagree.

Duplicate the instrument and **send it to all youth teachers** in Tarrant Association. Each teacher will read the statements and mark his or her degree of agreement (or disagreement) with each statement. Completed forms will be returned to the researcher by means of self-addressed and stamped envelopes.

Score forms just like a Likert scale. Scores for each statement produces a mean for the entire group. Means (and their associated statements) will then be **ranked**.

From this ranking, the researcher can determine how the group responded to the "major concerns" submitted by individuals earlier. **These will either be reinforced by agreement by the entire group (major concerns, indeed!), or they will be identified as**

¹Procedure described by Dr. John Curry, University of North Texas, EDER 601, Fall 1983

a isolated concerns not shared by the group.

The Delphi Technique is a powerful way to allow a group of subjects to create their own attitude statements, and then measure the strength (or lack) of support by the whole group for the statements generated by the process.¹

Summary

In this chapter we have introduced ways researchers measure attitudes. We have emphasized the Likert and Thurstone scales, the Q-Sort, and the Semantic Differential. These are but a sampling of procedures available to you to measure the subjective characteristics of groups.

Vocabulary

Evaluative	A scale in the semantic differential which measures good-bad
Likert scale	Attitude scale which uses + and - equally weighted statements
Potency	A scale in the semantic differential which measures strong-weak
Q-sort	Method for rank ordering a large number of attitudinal statements
Activity	A scale in the semantic differential which measures fast-slow
Semantic Differential	An attitude scale which measures differences in word meanings
Thurstone scale	Attitude scale which uses weighted statements

Study Questions

1. Define "attitude scale."
2. Compare and contrast the Likert and Thurston attitude scales.
3. What applications would be appropriate for the semantic differential in Christian research? Likert scale? Thurstone scale? Delphi Technique?

Sample Test Questions

1. The attitude scaling technique which uses equally weighted items is the
 - A. Likert Scale
 - B. Thurstone Scale
 - C. Q-Sort
 - D. Semantic Differential
2. The best approach to rank ordering a large number of statements is the
 - A. Likert Scale
 - B. Thurstone Scale
 - C. Q-Sort
 - D. Semantic Differential
3. The method to use in measuring the differences between selected groups in the way they use specified terms is the
 - A. Likert Scale
 - B. Thurstone Scale
 - C. Q-Sort
 - D. Semantic Differential

Sample Thurstone Scale

Instructions: Read the statements below, and check off any that reflects your attitude toward education. You may check off as many as you like. (weights on next page)

- I am intensely interested in education.
- I go to school only because I am compelled to do so.
- I am interested in education but one shouldn't get too concerned about it.
- I like reading thrillers and playing games better than studying.
- Education is of first rate importance in the life of man.
- Sometimes I feel education is necessary and sometimes I doubt it.
- I wouldn't work at studying so hard if I didn't have to pass exams.
- Education tends to make people snobs.
- I think time spent studying is wasted.
- It is better to start a career at age 18 than to go to college.
- It is doubtful that education has helped the world.
- I have no desire to have anything to do with education.
- We cannot become good citizens unless we are educated.
- More money should be spent on education.
- I think my education will be of use to me after I leave school.
- I always read newspaper articles on education.
- Education does more harm than good.
- I see no value in education.
- Education allows us to live a less monotonous life.
- I dislike education because that time has to be spent on homework.
- I like the subjects taught in school but do not like attending school.
- Education is doing more harm than good.
- Lack of education is the source of all evil.
- Education enables us to make the best possible use of our lives.
- Only educated people can enjoy life to the full.
- Education does more good than harm.
- I do not like school teachers so I somewhat dislike education.
- Education is alright in moderation.
- It is enough that we should be taught to read, write and do sums.
- I do not care about education so long as I can live comfortably.
- Education makes people forget God and despise Christianity.
- Education is an excellent character builder.
- Too much money is spent on education.
- If anything, I must admit to a slight dislike of education.

Attitude score = $1.0 + 1.3 + 2.7 + 1.8/4 = 1.3$ Very positive!

Sample Thurstone Scale (with weights)

Subject score equals average of weights of statements selected.

- 1.0 I am intensely interested in education.
- 10.0 I go to school only because I am compelled to do so.
- 4.2 I am interested in education but one shouldn't get too concerned about it.
- 6.4 I like reading thrillers and playing games better than studying.
- 0.5 Education is of first rate importance in the life of man.
- 5.4 Sometimes I feel education is necessary and sometimes I doubt it.
- 6.9 I wouldn't work at studying so hard if I didn't have to pass exams.
- 8.4 Education tends to make people snobs.
- 10.1 I think time spent studying is wasted.
- 7.9 It is better to start a career at age 18 than to go to college.
- 5.7 It is doubtful that education has helped the world.
- 10.9 I have no desire to have anything to do with education.
- 1.3 We cannot become good citizens unless we are educated.
- 2.2 More money should be spent on education.
- 3.7 I think my education will be of use to me after I leave school.
- 3.0 I always read newspaper articles on education.
- 9.3 Education does more harm than good.
- 11.4 I see no value in education.
- 3.3 Education allows us to live a less monotonous life.
- 7.4 I dislike education because time has to be spent on homework.
- 4.5 I like the subjects taught in school but do not like attending school.
- 10.5 Education is doing more harm than good.
- 2.3 Lack of education is the source of all evil.
- 0.3 Education enables us to make the best possible use of our lives.
- 1.2 Only educated people can enjoy life to the full.
- 2.7 Education does more good than harm.
- 7.1 I do not like school teachers so I somewhat dislike education.
- 4.9 Education is alright in moderation.
- 5.8 It is enough that we should be taught to read, write and do sums.
- 8.9 I do not care about education so long as I can live comfortably.
- 9.9 Education makes people forget God and despise Christianity.
- 1.8 Education is an excellent character builder.
- 8.6 Too much money is spent on education.
- 6.7 If anything, I must admit to a slight dislike of education.

