

General Issues in Scaling

[[Home](#)] [[General Issues in Scaling](#)] [[Thurstone Scaling](#)] [[Likert Scaling](#)] [[Guttman Scaling](#)]

S.S. Stevens came up with what I think is the simplest and most straightforward definition of scaling. He said:

Scaling is the assignment of objects to numbers according to a rule.

But what does that mean? In most scaling, the objects are text statements, usually statements of attitude or belief. The figure shows an example. There are three statements describing attitudes towards immigration. To scale these statements, we have to

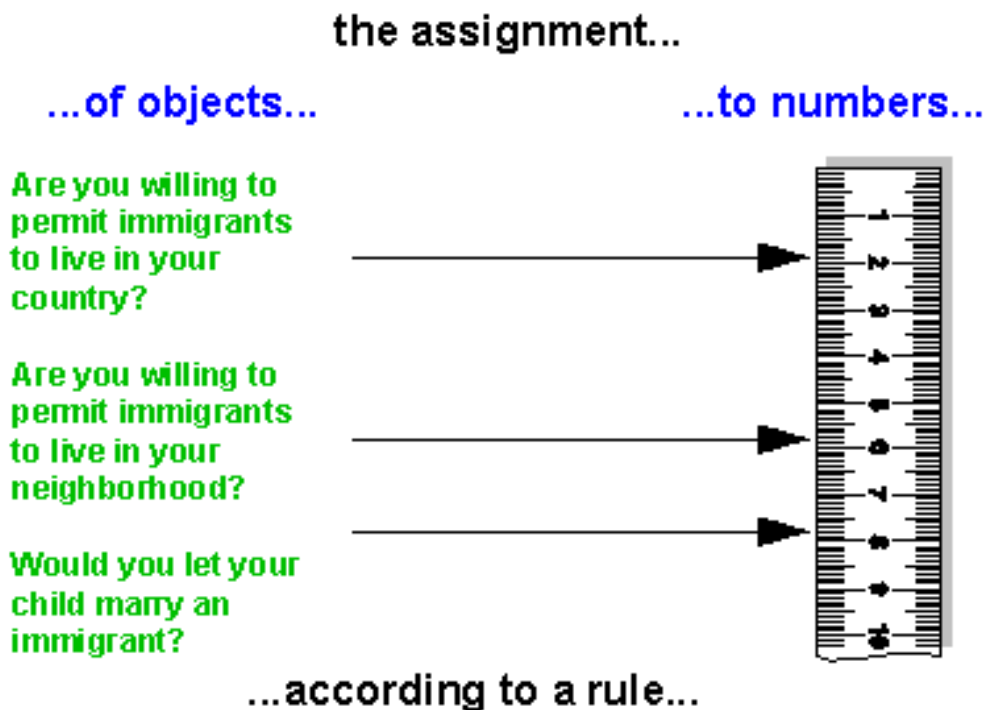
assign numbers to them.

Usually, we would like the result to be on at least an interval scale (see [Levels of Measurement](#)) as indicated by the ruler in the figure.

And what does "according to a rule" mean? If you look at the statements, you can see that as you read down, the attitude towards immigration becomes more restrictive -- if a person

agrees with a statement on the list, it's likely that they will also agree with all of the statements higher on the list. In this case, the "rule" is a *cumulative* one. So what is scaling? It's how we get numbers that can be meaningfully assigned to objects -- it's a set of procedures. We'll present several different approaches below.

But first, I have to clear up one of my pet peeves. People often confuse the idea of a scale and a response scale. A response scale is the way you collect responses from people on an instrument. You might use a dichotomous response scale like Agree/Disagree, True/False, or Yes/No. Or, you might use an interval response scale like a 1-to-5 or 1-to-7 rating. But, if all you are doing is attaching a response scale to an object or statement, you can't call that scaling. As you will see, scaling involves procedures that you do independent of the



respondent so that you can come up with a numerical value for the object. In true scaling research, you use a scaling procedure to develop your instrument (scale) and you also use a response scale to collect the responses from participants. But just assigning a 1-to-5 response scale for an item is **not** scaling! The differences are illustrated in the table below.

Scale	Response Scale
results from a process	is used to collect the response for an item
each item on scale has a scale value	item not associated with a scale value
refers to a set of items	used for a single item

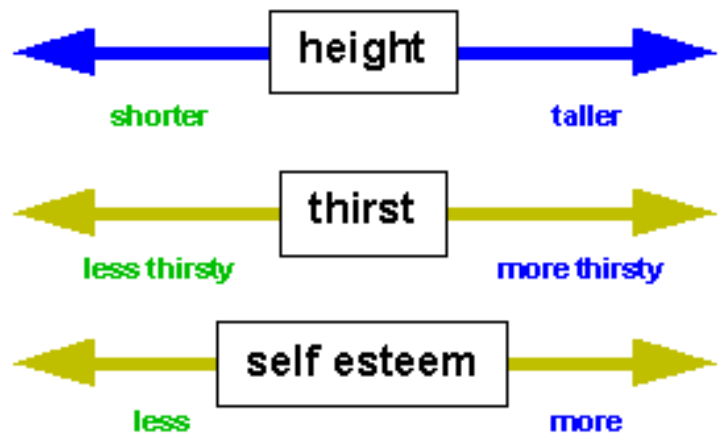
Purposes of Scaling

Why do we do scaling? Why not just create text statements or questions and use response formats to collect the answers? First, sometimes we do scaling to test a hypothesis. We might want to know whether the construct or concept is a single dimensional or multidimensional one (more about dimensionality later). Sometimes, we do scaling as part of exploratory research. We want to know what dimensions underlie a set of ratings. For instance, if you create a set of questions, you can use scaling to determine how well they "hang together" and whether they measure one concept or multiple concepts. But probably the most common reason for doing scaling is for scoring purposes. When a participant gives their responses to a set of items, we often would like to assign a single number that represents that's person's overall attitude or belief. For the figure above, we would like to be able to give a single number that describes a person's attitudes towards immigration, for example.

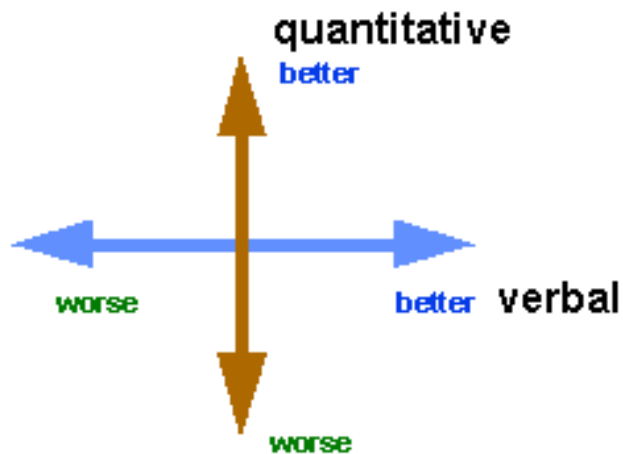
Dimensionality

A scale can have any number of dimensions in it. Most scales that we develop have only a few dimensions. What's a dimension? Think of a dimension as a number line. If we want to measure a construct, we have to decide whether the construct can be measured well with one number line or whether it may need more. For instance, height is a concept that is

unidimensional or one-dimensional. We can measure the concept of height very well with only a single number line (e.g., a ruler). Weight is also unidimensional -- we can measure it with a scale. Thirst might also be considered a unidimensional concept -- you are either more or less thirsty at any given time. It's easy to see that height and weight are unidimensional. But what about a concept like self esteem? If you think you can measure a person's self esteem well with a single ruler that goes from low to high, then you probably have a unidimensional construct.

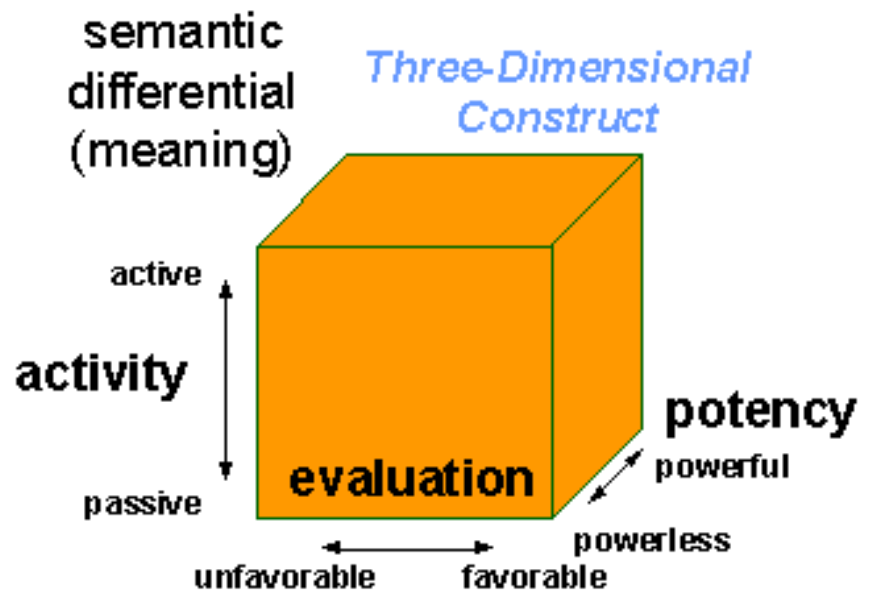


What would a two-dimensional concept be? Many models of intelligence or achievement postulate two major dimensions -- mathematical and verbal ability. In this type of two-dimensional model, a person can be said to possess two types of achievement. Some people will be high in verbal skills and lower in math. For others, it will be the reverse. But, if a concept is truly two-dimensional, it is not possible to depict a person's level on it using only a single number line. In other words, in order to describe achievement you would need to locate a person as a point in two dimensional (x,y) space.



OK, let's push this one step further: how about a three-dimensional concept? Psychologists who study the idea of meaning theorized that the meaning of a term could be well described in three dimensions. Put in other terms, any objects can be distinguished or differentiated from each other along three dimensions. They labeled these three dimensions *activity*, *evaluation*, and *potency*. They called this general theory of meaning the **semantic differential**. Their theory essentially states that you can rate any object along those three

dimensions. For instance, think of the idea of "ballet." If you like the ballet, you would probably rate it high on activity, favorable on evaluation, and powerful on potency. On the other hand, think about the concept of a "book" like a novel. You might rate it low on activity (it's passive), favorable on evaluation (assuming you like it), and about average on potency. Now, think of the idea of "going to the dentist." Most people would rate it low on activity (it's a passive activity), unfavorable on evaluation, and powerless on potency (there are few routine activities that make you feel as powerless!). The theorists who came up with the idea of the semantic differential thought that the meaning of any concepts could be described well by rating the concept on these three dimensions. In other words, in order to describe the meaning of an object you have to locate it as a dot somewhere within the cube (three-dimensional space).



Unidimensional or Multidimensional?

What are the advantages of using a unidimensional model? Unidimensional concepts are generally easier to understand. You have either more or less of it, and that's all. You're either taller or shorter, heavier or lighter. It's also important to understand what a unidimensional scale is as a foundation for comprehending the more complex multidimensional concepts. But the best reason to use unidimensional scaling is because you believe the concept you are measuring really is unidimensional in reality. As you've seen, many familiar concepts (height, weight, temperature) are actually unidimensional. But, if the concept you are studying is in fact multidimensional in nature, a unidimensional scale or number line won't describe it well. If you try to measure academic achievement on a single dimension, you would place every person on a single line ranging from low to high achievers. But how do you score someone who is a high math achiever and terrible verbally, or vice versa? A unidimensional scale can't capture that type of achievement.

The Major Unidimensional Scale Types

There are three major types of unidimensional scaling methods. They are similar in that they each measure the concept of interest on a number line. But they differ considerably in how they arrive at scale values for different items. The three methods are Thurstone or Equal-Appearing Interval Scaling, Likert or "Summative" Scaling, and Guttman or

"Cumulative" Scaling.

Copyright ©2002, William M.K. Trochim, All Rights Reserved
[Purchase a printed copy of the Research Methods Knowledge Base](#)

Thurstone Scaling

[[Home](#)] [[General Issues in Scaling](#)] [[Thurstone Scaling](#)] [[Likert Scaling](#)] [[Guttman Scaling](#)]

Thurstone was one of the first and most productive scaling theorists. He actually invented three different methods for developing a unidimensional scale: the **method of equal-appearing intervals**; the **method of successive intervals**; and, the **method of paired comparisons**. The three methods differed in how the scale values for items were constructed, but in all three cases, the resulting scale was rated the same way by respondents. To illustrate Thurstone's approach, I'll show you the easiest method of the three to implement, the method of equal-appearing intervals.

The Method of Equal-Appearing Intervals

Developing the Focus. The Method of Equal-Appearing Intervals starts like almost every other scaling method -- with a large set of statements. Oops! I did it again! You can't start with the set of statements -- you have to first define the focus for the scale you're trying to develop. Let this be a warning to all of you: methodologists like me often start our descriptions with the first objective methodological step (in this case, developing a set of statements) and forget to mention critical foundational issues like the development of the focus for a project. So, let's try this again...

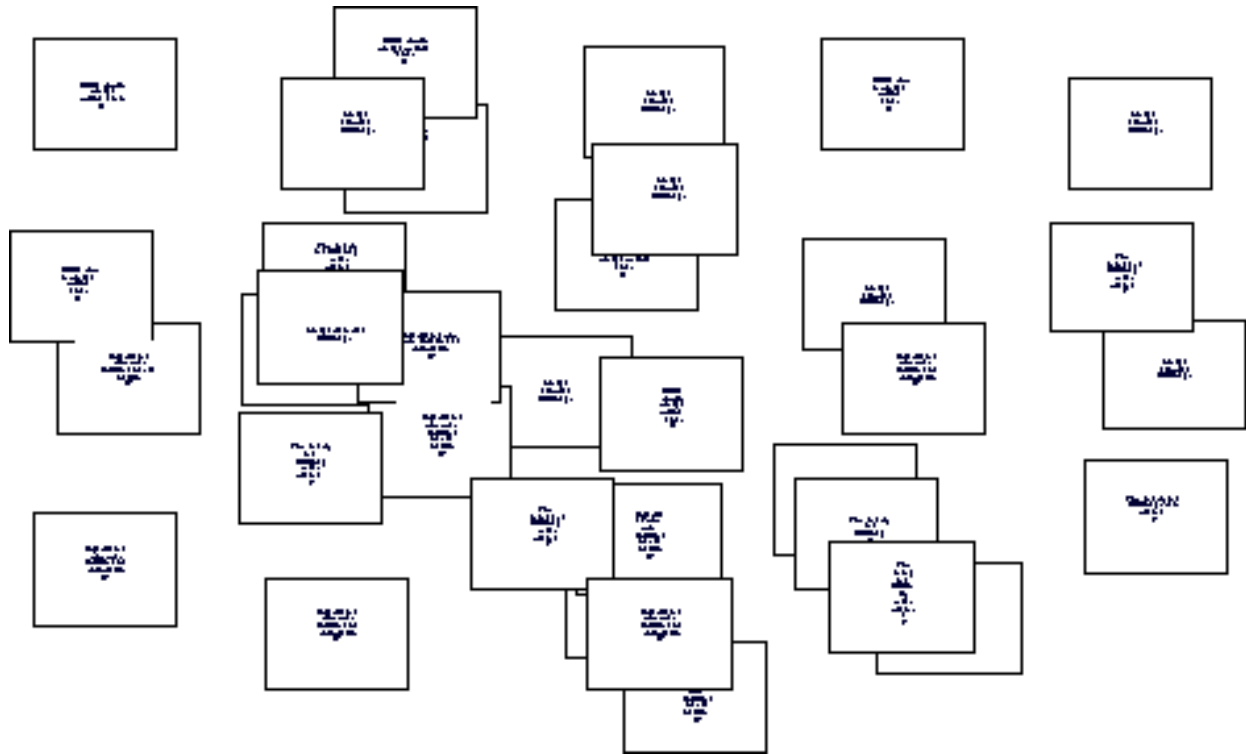
The Method of Equal-Appearing Intervals starts like almost every other scaling method -- with the development of the focus for the scaling project. Because this is a unidimensional scaling method, we assume that the concept you are trying to scale is reasonably thought of as one-dimensional. The description of this concept should be as clear as possible so that the person(s) who are going to create the statements have a clear idea of what you are trying to measure. I like to state the focus for a scaling project in the form of a command -- the command you will give to the people who will create the statements. For instance, you might start with the focus command:

Generate statements that describe specific attitudes that people might have towards persons with AIDS.

You want to be sure that everyone who is generating statements has some idea of what you are after in this focus command. You especially want to be sure that technical language and acronyms are spelled out and understood (e.g., what is AIDS?).














Generating Potential Scale Items.



































Now, you're ready to create statements. You want a large set of candidate statements (e.g., 80 -- 100) because you are going to select your final scale items from this pool. You also



want to be sure that all of the statements are worded similarly -- that they don't differ in grammar or structure. For instance, you might want them each to be worded as a statement which you could agree or disagree with. You don't want some of them to be statements while others are questions.

For our example focus on developing an AIDS attitude scale, we might generate statements like the following (these statements came from a class exercise I did in my Spring 1997 undergrad class):

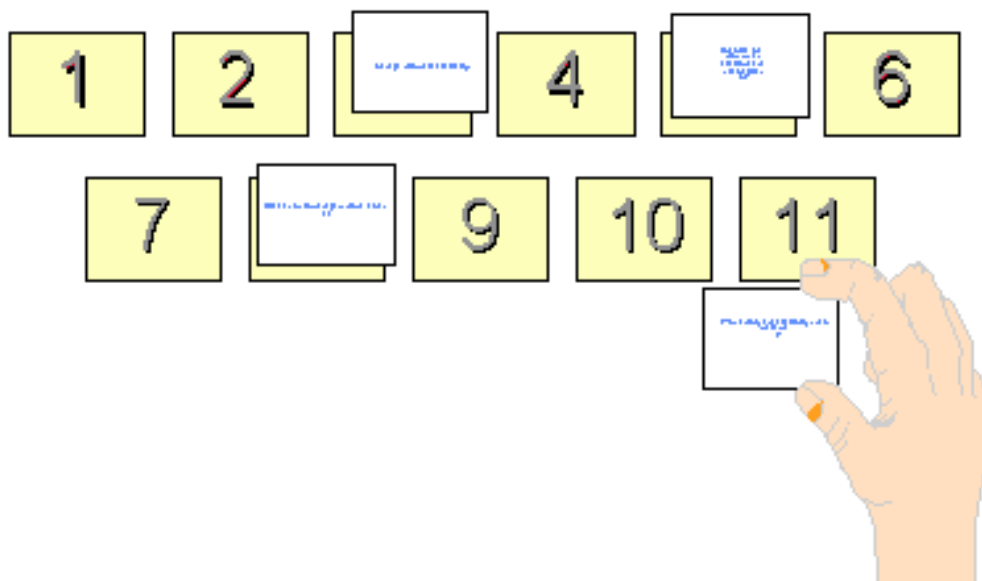
-  people get AIDS by engaging in immoral behavior
-  you can get AIDS from toilet seats
-  AIDS is the wrath of God
-  anybody with AIDS is either gay or a junkie
-  AIDS is an epidemic that affects us all
-  people with AIDS are bad
-  people with AIDS are real people
-  AIDS is a cure, not a disease
-  you can get AIDS from heterosexual sex
-  people with AIDS are like my parents
-  you can get AIDS from public toilets
-  women don't get AIDS
-  I treat everyone the same, regardless of whether or not they have AIDS
-  AIDS costs the public too much

-  AIDS is something the other guy gets
-  living with AIDS is impossible
-  children cannot catch AIDS
-  AIDS is a death sentence
-  because AIDS is preventable, we should focus our resources on prevention instead of curing
-  People who contract AIDS deserve it
-  AIDS doesn't have a preference, anyone can get it.
-  AIDS is the worst thing that could happen to you.
-  AIDS is good because it will help control the population.
-  If you have AIDS, you can still live a normal life.
-  People with AIDS do not need or deserve our help
-  By the time I would get sick from AIDS, there will be a cure
-  AIDS will never happen to me
-  you can't get AIDS from oral sex
-  AIDS is spread the same way colds are
-  AIDS does not discriminate
-  You can get AIDS from kissing
-  AIDS is spread through the air
-  Condoms will always prevent the spread of AIDS
-  People with AIDS deserve what they got
-  If you get AIDS you will die within a year
-  Bad people get AIDS and since I am a good person I will never get AIDS
-  I don't care if I get AIDS because researchers will soon find a cure for it.
-  AIDS distracts from other diseases that deserve our attention more
-  bringing AIDS into my family would be the worst thing I could do
-  very few people have AIDS, so it's unlikely that I'll ever come into contact with a sufferer
-  if my brother caught AIDS I'd never talk to him again
-  People with AIDS deserve our understanding, but not necessarily special treatment
-  AIDS is a omnipresent, ruthless killer that lurks around dark alleys, silently waiting for naive victims to wander passed so that it might pounce.
-  I can't get AIDS if I'm in a monogamous relationship
-  the nation's blood supply is safe
-  universal precautions are infallible
-  people with AIDS should be quarantined to protect the rest of society
-  because I don't live in a big city, the threat of AIDS is very small

- I know enough about the spread of the disease that I would have no problem working in a health care setting with patients with AIDS
- the AIDS virus will not ever affect me
- Everyone affected with AIDS deserves it due to their lifestyle
- Someone with AIDS could be just like me
- People infected with AIDS did not have safe sex
- Aids affects us all.
- People with AIDS should be treated just like everybody else.
- AIDS is a disease that anyone can get if there are not careful.
- It's easy to get AIDS.
- The likelihood of contracting AIDS is very low.
- The AIDS quilt is an emotional reminder to remember those who did not deserve to die painfully or in vain
- The number of individuals with AIDS in Hollywood is higher than the general public thinks
- It is not the AIDS virus that kills people, it is complications from other illnesses (because the immune system isn't functioning) that cause death
- AIDS is becoming more a problem for heterosexual women and their offsprings than IV drug users or homosexuals
- A cure for AIDS is on the horizon
- A cure for AIDS is on the horizon
- Mandatory HIV testing should be established for all pregnant women

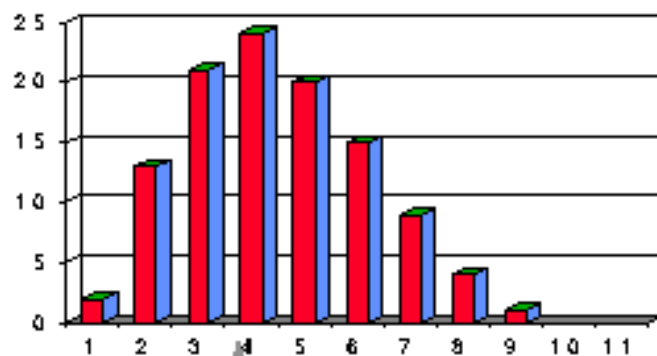
1 = least favorable to the concept

11 = most favorable to the concept



Rating the Scale Items. OK, so now you have a set of statements. The next step is to have your participants (i.e., judges) rate each statement on a 1-to-11 scale in terms of how much each statement indicates a **favorable** attitude towards people with AIDS. Pay close attention here! You DON'T want the participants to tell you what their attitudes towards AIDS are, or whether they would agree with the statements. You want them to rate the "favorableness" of each statement in terms of an attitude towards AIDS, where 1 = "extremely unfavorable attitude towards people with AIDS" and 11 = "extremely favorable attitude towards people with AIDS.". (Note that I could just as easily had the judges rate how much each statement represents a negative attitude towards AIDS. If I did, the scale I developed would have higher scale values for people with more negative attitudes).

For each item, plot the distribution of pile numbers...



get the median
and interquartile range

Computing Scale Score Values for Each Item. The next step is to analyze the rating data. For each statement, you need to compute the Median and the Interquartile Range. The median is the value above and below which 50% of the ratings fall. The first quartile (Q1) is the value below which 25% of the cases fall and above which 75% of the cases fall -- in other words, the 25th percentile. The median is the 50th percentile. The third quartile, Q3, is the 75th percentile. The Interquartile Range is the difference between third and first quartile, or $Q3 - Q1$. The figure above shows a histogram for a single item and indicates the median and Interquartile Range. You can compute these values easily with any introductory statistics program or with most spreadsheet programs. To facilitate the final selection of items for your scale, you might want to sort the table of medians and Interquartile Range in ascending order by Median and, within that, in descending order by Interquartile Range. For the items in this example, we got a table like the following:













Statement Number	Median	Q1	Q3	Interquartile Range
23	1	1	2.5	1.5
8	1	1	2	1
12	1	1	2	1
34	1	1	2	1
39	1	1	2	1
54	1	1	2	1
56	1	1	2	1
57	1	1	2	1
18	1	1	1	0
25	1	1	1	0
51	1	1	1	0
27	2	1	5	4
45	2	1	4	3
16	2	1	3.5	2.5
42	2	1	3.5	2.5
24	2	1	3	2
44	2	2	4	2
36	2	1	2.5	1.5
43	2	1	2.5	1.5
33	3	1	5	4
48	3	1	5	4
20	3	1.5	5	3.5
28	3	1.5	5	3.5
31	3	1.5	5	3.5
19	3	1	4	3
22	3	1	4	3
37	3	1	4	3
41	3	2	5	3
6	3	1.5	4	2.5
21	3	1.5	4	2.5

32	3	2	4.5	2.5
9	3	2	3.5	1.5
1	4	3	7	4
26	4	1	5	4
47	4	1	5	4
30	4	1.5	5	3.5
13	4	2	5	3
11	4	2	4.5	2.5
15	4	3	5	2
40	5	4.5	8	3.5
2	5	4	6.5	2.5
14	5	4	6	2
17	5.5	4	8	4
49	6	5	9.75	4.75
50	8	5.5	11	5.5
35	8	6.25	10	3.75
29	9	5.5	11	5.5
38	9	5.5	10.5	5
3	9	6	10	4
55	9	7	11	4
10	10	6	10.5	4.5
7	10	7.5	11	3.5
46	10	8	11	3
5	10	8.5	11	2.5
53	11	9.5	11	1.5
4	11	10	11	1

Selecting the Final Scale Items. Now, you have to select the final statements for your scale. You should select statements that are at equal intervals across the range of medians. In our example, we might select one statement for each of the eleven median values. Within each value, you should try to select the statement that has the smallest Interquartile Range. This is the statement with the least amount of variability across judges. You don't want the statistical analysis to be the only deciding factor here. Look over the candidate statements at each level and select the statement that makes the most sense. If you find that the best

statistical choice is a confusing statement, select the next best choice.

When we went through our statements, we came up with the following set of items for our scale:

-  People with AIDS are like my parents (6)
-  Because AIDS is preventable, we should focus our resources on prevention instead of curing (5)
-  People with AIDS deserve what they got. (1)
-  Aids affects us all (10)
-  People with AIDS should be treated just like everybody else. (11)
-  AIDS will never happen to me. (3)
-  It's easy to get AIDS (5)
-  AIDS doesn't have a preference, anyone can get it (9)
-  AIDS is a disease that anyone can get if they are not careful (9)
-  If you have AIDS, you can still lead a normal life (8)
-  AIDS is good because it helps control the population. (2)
-  I can't get AIDS if I'm in a monogamous relationship. (4)

The value in parentheses after each statement is its scale value. Items with higher scale values should, in general, indicate a more favorable attitude towards people with AIDS. Notice that we have randomly scrambled the order of the statements with respect to scale values. Also, notice that we do not have an item with scale value of 7 and that we have two with values of 9.

Administering the Scale. You now have a scale -- a yardstick you can use for measuring attitudes towards people with AIDS. You can give it to a participant and ask them to agree or disagree with each statement. To get that person's total scale score, you average the scale scores of all the items that person agreed with. For instance, let's say a respondent completed the scale as follows:

Agree	Disagree	People with AIDS are like my parents.
-------	----------	---------------------------------------

Agree	Disagree	Because AIDS is preventable, we should focus our resources on prevention instead of curing.
Agree	Disagree	People with AIDS deserve what they got.
Agree	Disagree	Aids affects us all.
Agree	Disagree	People with AIDS should be treated just like everybody else.
Agree	Disagree	AIDS will never happen to me.
Agree	Disagree	It's easy to get AIDS.
Agree	Disagree	AIDS doesn't have a preference, anyone can get it.
Agree	Disagree	AIDS is a disease that anyone can get if they are not careful.
Agree	Disagree	If you have AIDS, you can still lead a normal life.

Agree	Disagree	AIDS is good because it helps control the population.
Agree	Disagree	I can't get AIDS if I'm in a monogamous relationship.

If you're following along with the example, you should see that the respondent checked eight items as Agree. When we take the average scale values for these eight items, we get a final value for this respondent of 7.75. This is where this particular respondent would fall on our "yardstick" that measures attitudes towards persons with AIDS. Now, let's look at the responses for another individual:

Agree	Disagree	People with AIDS are like my parents.
Agree	Disagree	Because AIDS is preventable, we should focus our resources on prevention instead of curing.
Agree	Disagree	People with AIDS deserve what they got.
Agree	Disagree	Aids affects us all.

Agree	Disagree	People with AIDS should be treated just like everybody else.
Agree	Disagree	AIDS will never happen to me.
Agree	Disagree	It's easy to get AIDS.
Agree	Disagree	AIDS doesn't have a preference, anyone can get it.
Agree	Disagree	AIDS is a disease that anyone can get if they are not careful.
Agree	Disagree	If you have AIDS, you can still lead a normal life.
Agree	Disagree	AIDS is good because it helps control the population.
Agree	Disagree	I can't get AIDS if I'm in a monogamous relationship.

In this example, the respondent only checked four items, all of which are on the negative end of the scale. When we average the scale items for the statements with which the respondent agreed we get an average score of 2.5, considerably lower or more negative in

attitude than the first respondent.

The Other Thurstone Methods

The other Thurstone scaling methods are similar to the Method of Equal-Appearing Intervals. All of them begin by focusing on a concept that is assumed to be unidimensional and involve generating a large set of potential scale items. All of them result in a scale consisting of relatively few items which the respondent rates on Agree/Disagree basis. The major differences are in how the data from the judges is collected. For instance, the method of paired comparisons requires each judge to make a judgement about each pair of statements. With lots of statements, this can become very time consuming indeed. With 57 statements in the original set, there are 1,596 unique pairs of statements that would have to be compared! Clearly, the paired comparison method would be too time consuming when there are lots of statements initially.

Thurstone methods illustrate well how a simple unidimensional scale might be constructed. There are other approaches, most notably [Likert or Summative Scales](#) and [Guttman or Cumulative Scales](#).

Copyright ©2002, William M.K. Trochim, All Rights Reserved
[Purchase a printed copy of the Research Methods Knowledge Base](#)

Likert Scaling

[[Home](#)] [[General Issues in Scaling](#)] [[Thurstone Scaling](#)] [[Likert Scaling](#)] [[Guttman Scaling](#)]

Like [Thurstone](#) or [Guttman](#) Scaling, Likert Scaling is a [unidimensional scaling](#) method. Here, I'll explain the basic steps in developing a Likert or "Summative" scale.

Defining the Focus. As in all scaling methods, the first step is to define what it is you are trying to measure. Because this is a unidimensional scaling method, it is assumed that the concept you want to measure is one-dimensional in nature. You might operationalize the definition as an instruction to the people who are going to create or generate the initial set of candidate items for your scale.

Generating the Items.

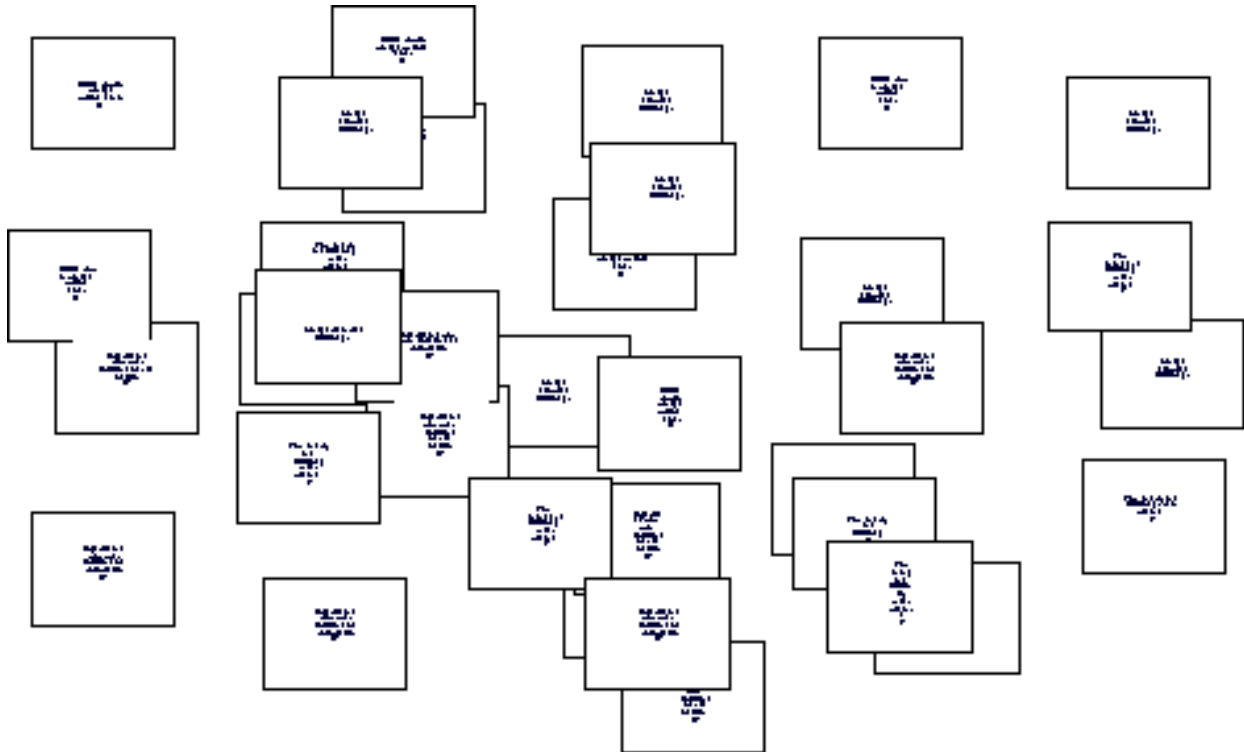
next, you have to create the set of potential scale items.

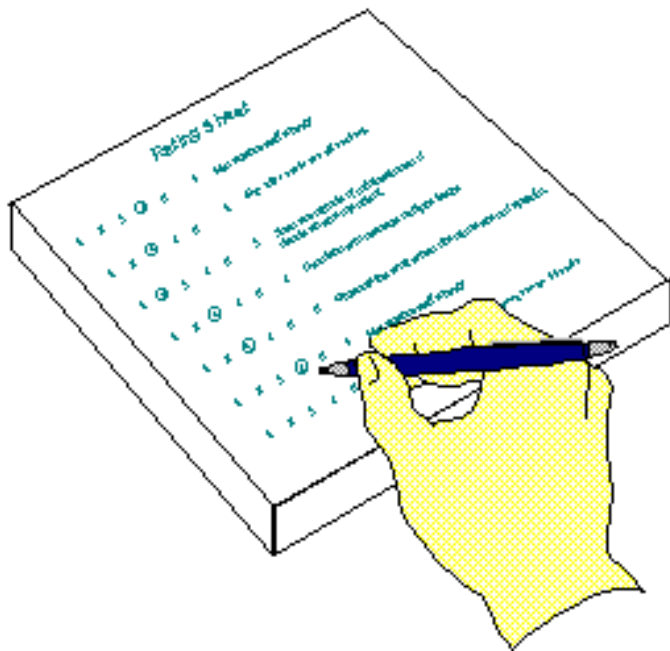
These should be items that can be rated on a 1-to-5 or 1-to-7

Disagree-Agree response scale.

Sometimes you can create

the items by yourself based on your intimate understanding of the subject matter. But, more often than not, it's helpful to engage a number of people in the item creation step. For instance, you might use some form of brainstorming to create the items. It's desirable to have as large a set of potential items as possible at this stage, about 80-100 would be best.





Rating the Items. The next step is to have a group of judges rate the items. Usually you would use a 1-to-5 rating scale where:

1. = strongly unfavorable to the concept
2. = somewhat unfavorable to the concept
3. = undecided
4. = somewhat favorable to the concept
5. = strongly favorable to the concept

Notice that, as in other scaling methods, the judges are not telling you what they believe -- they are judging how favorable each item is with respect to the construct of interest.

Selecting the Items. The next step is to compute the intercorrelations between all pairs of items, based on the ratings of the judges. In making judgements about which items to retain for the final scale there are several analyses you can do:



- Throw out any items that have a low correlation with the total (summed) score across all items

In most statistics packages it is relatively easy to compute this type of Item-Total correlation. First, you create a new variable which is the sum of all of the individual items for each respondent. Then, you include this variable in the correlation matrix computation (if you include it as the last variable in the list, the resulting Item-Total correlations will all be the last line of the correlation matrix and will be easy to spot). How low should the correlation be for you to throw out the item? There is no fixed rule here -- you might eliminate all items with a correlation with the total score less than .6, for example.



For each item, get the average rating for the top quarter of judges and the bottom quarter. Then, do a t-test of the differences between the mean value for the item for the top and bottom quarter judges.

Higher t-values mean that there is a greater difference between the highest and lowest judges. In more practical terms, items with higher t-values are better discriminators, so you want to keep these items. In the end, you will have to use your judgement about which items are most sensibly retained. You want a relatively small number of items on your final scale (e.g., 10-15) and you want them to have high Item-Total correlations and high discrimination (e.g., high t-values).

Administering the Scale. You're now ready to use your Likert scale. Each respondent is asked to rate each item on some response scale. For instance, they could rate each item on a 1-to-5 response scale where:

1. = strongly disagree
2. = disagree
3. = undecided
4. = agree
5. = strongly agree

There are a variety possible response scales (1-to-7, 1-to-9, 0-to-4). All of these odd-numbered scales have a middle value is often labeled Neutral or Undecided. It is also possible to use a forced-choice response scale with an even number of responses and no middle neutral or undecided choice. In this situation, the respondent is forced to decide whether they lean more towards the agree or disagree end of the scale for each item.

The final score for the respondent on the scale is the sum of their ratings for all of the items (this is why this is sometimes called a "summated" scale). On some scales, you will have items that are reversed in meaning from the overall direction of the scale. These are called **reversal items**. You will need to reverse the response value for each of these items before summing for the total. That is, if the respondent gave a 1, you make it a 5; if they gave a 2 you make it a 4; 3 = 3; 4 = 2; and, 5 = 1.

Example: The Employment Self Esteem Scale

Here's an example of a ten-item Likert Scale that attempts to estimate the level of self esteem a person has on the job. Notice that this instrument has no center or neutral point -- the respondent has to declare whether he/she is in agreement or disagreement with the item.

INSTRUCTIONS: Please rate how strongly you agree or disagree with each of the following statements by placing a check mark in the appropriate box.

Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	1. I feel good about my work on the job.
Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	2. On the whole, I get along well with others at work.
Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	3. I am proud of my ability to cope with difficulties at work.
Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	4. When I feel uncomfortable at work, I know how to handle it.
Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	5. I can tell that other people at work are glad to have me there.
Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	6. I know I'll be able to cope with work for as long as I want.
Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	7. I am proud of my relationship with my supervisor at work.
Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	8. I am confident that I can handle my job without constant assistance.
Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	9. I feel like I make a useful contribution at work.
Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	10. I can tell that my coworkers respect me.







Guttman Scaling

[[Home](#)] [[General Issues in Scaling](#)] [[Thurstone Scaling](#)] [[Likert Scaling](#)] [[Guttman Scaling](#)]

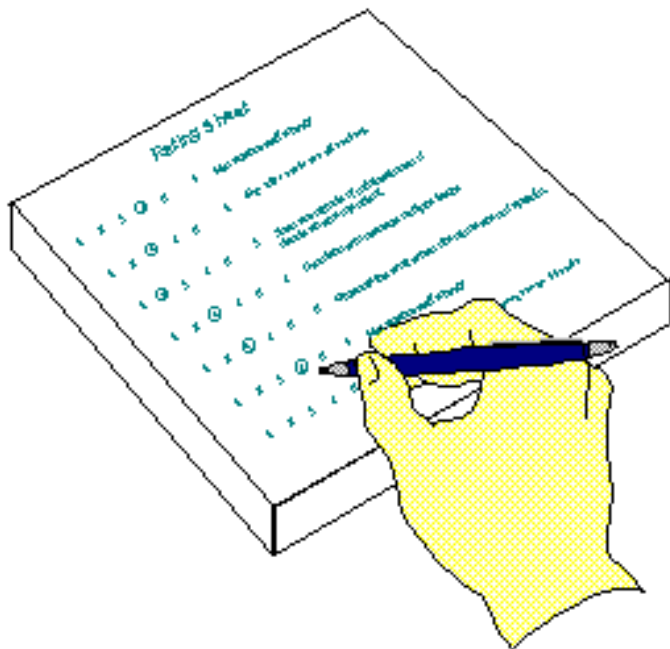
Guttman scaling is also sometimes known as **cumulative scaling** or **scalogram analysis**. The purpose of Guttman scaling is to establish a one-dimensional continuum for a concept you wish to measure. What does that mean? Essentially, we would like a set of items or statements so that a respondent who agrees with any specific question in the list will also agree with all previous questions. Put more formally, we would like to be able to predict item responses perfectly knowing only the total score for the respondent. For example, imagine a ten-item cumulative scale. If the respondent scores a four, it should mean that he/she agreed with the first four statements. If the respondent scores an eight, it should mean they agreed with the first eight. The object is to find a set of items that perfectly matches this pattern. In practice, we would seldom expect to find this cumulative pattern perfectly. So, we use scalogram analysis to examine how closely a set of items corresponds with this idea of cumulativeness. Here, I'll explain how we develop a Guttman scale.

Define the Focus. As in all of the scaling methods, we begin by defining the focus for our scale. Let's imagine that you wish to develop a cumulative scale that measures U.S. citizen attitudes towards immigration. You would want to be sure to specify in your definition whether you are talking about any type of immigration (legal and illegal) from anywhere (Europe, Asia, Latin and South America, Africa).

Develop the Items. Next, as in all scaling methods, you would develop a large set of items that reflect the concept. You might do this yourself or you might engage a knowledgeable group to help. Let's say you came up with the following statements:

-  I would permit a child of mine to marry an immigrant.
-  I believe that this country should allow more immigrants in.
-  I would be comfortable if a new immigrant moved next door to me.
-  I would be comfortable with new immigrants moving into my community.
-  It would be fine with me if new immigrants moved onto my block.
-  I would be comfortable if my child dated a new immigrant.

Of course, we would want to come up with many more statements (about 80-100 would be desirable).



Rate the Items. Next, we would want to have a group of judges rate the statements or items in terms of how favorable they are to the concept of immigration. They would give a Yes if the item was favorable toward immigration and a No if it is not. Notice that we are not asking the judges whether they personally agree with the statement. Instead, we're asking them to make a judgment about how the statement is related to the construct of interest.

when sorted by row and column
will show whether there is
a **cumulative scale**







Develop the Cumulative Scale. The key to Guttman scaling is in the analysis. We construct a matrix or table that shows the responses of all the respondents on all of the items. We then sort this matrix so that respondents who agree with more statements are listed at the top and those agreeing with fewer are at the bottom. For respondents with the same number of agreements, we sort the statements from left to right from those that most agreed to to those that fewest agreed to. We might get a table something like the figure. Notice that the scale is very nearly cumulative when you read from left to right across the columns (items).

Specifically if someone agreed with Item 7, they always agreed with Item 2. And, if someone agreed with Item 5, they always agreed with Items 7 and 2. The matrix shows that the cumulateness of the scale is not perfect, however. While in general, a person agreeing with Item 3 tended to also agree with 5, 7 and 2, there are several exceptions to that rule.

While we can examine the matrix if there are only a few items in it, if there are lots of items, we need to use a data analysis called **scalogram analysis** to determine the subsets of items from our pool that best approximate the cumulative property. Then, we review these items and select our final scale elements. There are several statistical techniques for examining the table to find a cumulative scale. Because there is seldom a perfectly cumulative scale we usually have to test how good it is. These statistics also estimate a scale score value for each item. This scale score is used in the final calculation of a respondent's score.

Respondent	Item 2	Item 7	Item 5	Item 3
7	Y	Y	Y	Y
15	Y	Y	Y	—
3	Y	Y	Y	Y
29	Y	Y	Y	Y
19	Y	Y	Y	—
32	Y	Y	—	Y
41	Y	Y	—	—
6	Y	Y	—	—
14	Y	—	—	Y
33	—	—	—	—

Administering the Scale. Once you've selected the final scale items, it's relatively simple to administer the scale. You simply present the items and ask the respondent to check items with which they agree. For our hypothetical immigration scale, the items might be listed in cumulative order as:

-  I believe that this country should allow more immigrants in.
-  I would be comfortable with new immigrants moving into my community.
-  It would be fine with me if new immigrants moved onto my block.
-  I would be comfortable if a new immigrant moved next door to me.
-  I would be comfortable if my child dated a new immigrant.
-  I would permit a child of mine to marry an immigrant.

Of course, when we give the items to the respondent, we would probably want to mix up the order. Our final scale might look like:

INSTRUCTIONS: Place a check next to each statement you agree with.

_____ I would permit a child of mine to marry an immigrant.

_____ I believe that this country should allow more immigrants in.

_____ I would be comfortable if a new immigrant moved next door to me.

_____ I would be comfortable with new immigrants moving into my community.

_____ It would be fine with me if new immigrants moved onto my block.

_____ I would be comfortable if my child dated a new immigrant.

Each scale item has a scale value associated with it (obtained from the scalogram analysis). To compute a respondent's scale score we simply sum the scale values of every

item they agree with. In our example, their final value should be an indication of their attitude towards immigration.

Copyright ©2002, William M.K. Trochim, All Rights Reserved
[Purchase a printed copy of the Research Methods Knowledge Base](#)