

Statistics Corner

Questions and answers about language testing statistics:

What issues affect Likert-scale questionnaire formats?

James Dean Brown (University of Hawai'i at Manoa)

QUESTION: Recently I came across a survey which attempted to evaluate student interest about a range classroom topics. Students were asked to rank their interest in various potential topics according to this scale:

10	if they felt a topic was very interesting
6	if they felt a topic was above average interest
4	if they felt a topic was below average interest
1	if they felt a topic was not worth studying in class

Please note that only four responses were permitted: 10, 6, 4, and 1. Is this an acceptable survey design? Should the scale reflect the number of permissible responses, rather than an arbitrary figure of 10?

ANSWER: To begin with, I find it difficult to answer your question because I do not know the context in which the scale was used. Nonetheless, I will try to respond, based on what I do know, because it affords me an opportunity to write a little bit about the problems of designing such scales. I hope you will find that information useful.

The general type of questionnaire item you refer to in your question is called a Likert-scale item (named after Rensis Likert, and pronounced with a short i as in it). Likert-scale items are most often used to investigate how respondents rate a series of statements by having them circle or otherwise mark numbered categories (for instance, 1 2 3 4 5). Likert-scale items are useful for gathering respondents' feelings, opinions, attitudes, etc. on any language-related topics. Typically, the numbered categories are on continuums like the following: very serious to not at all serious, very important to very unimportant, strongly like to strongly dislike, or strongly agree to strongly disagree. Two problems commonly arise when trying to use Likert-scale items: (a) you may encounter some students who prefer to "sit the fence" by always marking the most neutral possible answer, and (b) you may find it difficult to decide what kind of scale the data coming from such an item represents. I will address both of those issues by way of providing background for a direct answer to your question.

Students Who "Sit the Fence"

The first problem is dealing with those students who tend to "sit the fence" on Likert-scale items. Given the possibility of a neutral option (like the 3 for don't know in a five-point strongly agree to strongly disagree), such students will tend to take that neutral option. If you need to force respondents to express a definite opinion one way or the other, you may want to use an even number of options (say four options like, 1 2 3 4) from which they must choose either in the positive or negative direction. When using such four-option Likert-scale items at the University of Hawaii, I have found that most students will pick 2 or 3, but they are at least expressing some opinion, one way or the other. However, even so, I have found a few students so prone to selecting the neutral answer that they circle the space between the 2 and the 3. I have therefore had to code some of the answers as 2.5. Nonetheless, using an even number of options forced the majority of students to go one way or the other.

Unfortunately, by doing that I may have been forcing students to have an opinion who really did not. That of course is another facet of this problem that you must consider when you are deciding whether to use an even or odd number of options, that is, some students really do feel neutral, or have no opinion about a particular issue, and you may want to know that. In such a case, you will want to give the respondents an odd number of options with a neutral position in the middle, or offer 1 2 3 4 as choices along with another option that is no opinion. The call is yours, and what you decide will depend on the kinds of information you want to get from your questionnaire.

Deciding the Type of Scale

The second problem is that you may find it difficult to decide what kind of scale the data coming from such a Likert-scale item represents. Three scales of measurement are often described in books on statistical analyses of surveys: categorical, rank-ordered, and continuous.

Categorical scales (also called nominal scales) quantify by tallying up the number in each of two or more categories. For example, a group might be made up of 21 females and only 10 males. That information taken together is a nominal scale with two categories (female and male). Other variables with more than two categories, like nationality, first language background, educational background, etc., are all potential categorical scales if the number of people in each category is being tallied.

Rank-ordered scales (also called ordinal scales) quantify by giving each data point a rank. For example, the students in a class might be ranked from 1st to 30th in terms of their test scores. That, or any other such ranking, would be a rank-ordered scale. Thus, any variable for which ordinal numbers are being used (1st, 2nd, 3rd, 4th, etc.) is a rank-ordered scale.

Continuous scales (also sometimes separated into interval and ratio scales) quantify at equal intervals along some yardstick. Thus inches, feet, and yards are equal intervals along a real yardstick and represent a continuous scale. Similarly, we treat IQ scores, TOEFL scores, and even classroom test scores as points along a continuum of possible scores. Hence, they are continuous, too. One other characteristic of continuous scores is that calculating means and standard deviations makes sense (which is not true of categorical or rank-ordered scales). (For much more on scales, see Brown, in press, chapter 1; Brown, 1988, pp. 20-28; 1996, pp. 93-98; 1999, pp. 109-115; or Hatch & Lazaraton, 1991).

Unfortunately, many a novice teacher-researcher has trouble deciding whether Likert scales are categorical, rank-ordered, or continuous. Sometimes, Likert scales are treated as categorical scales. For example, a researcher might report that five people chose 1, sixteen selected 2, twenty-six preferred 3, fourteen decided on 4, and three picked 5. Other times, Likert scales are referred to as a rank-order. For instance, again using the example in the previous sentence, five people would be reported as ranking the statement 1st, sixteen ranking it 2nd, twenty-six ranking it 3rd, fourteen ranking it 4th, and three ranking it 5th. Still other times, Likert scales are analyzed as continuous, with each set of 1 2 3 4 5 treated as equal points along a continuum. In such cases, a mean and standard deviation is often reported for each of the Likert-scale questions. Using the same example that runs throughout this paragraph, the mean would be 2.91 and the standard deviation would be .98.

A Direct Answer to Your Question

The 1, 4, 6, 10 scale you referred to in your question at the top of this article is a strange scale indeed. The scale cannot be considered continuous because the points on the scale are not equally spaced. As you can see if you examine where they sit along the continuum from one to ten: 1 2 3 4 5 6 7 8 9 10. Clearly, there are two numbers between the 1 and 4, one number between the 4 and 6, and three numbers between the 6 and 10. Thus the numbers are not equally spaced along a continuum and therefore do not form a continuous scale. Similarly, these four numbers are not rank-ordered because they are not ordinal in nature (that is, saying 1st, 4th, 6th, and 10th would make absolutely no sense). At best this scale might be analyzed as categorical, but convincing readers (e.g., The "Confused Teacher" who inspired this column) that the categories make sense might be difficult (because they are not evenly spaced along a continuum). Thus, whoever the researcher was who used the 1 4 6 10 Likert-like scale took a scale which is already difficult to analyze and made it more difficult to deal with.

In short, the scale probably would have been better as a more traditional 1 2 3 4 5, or 1 2 3 4, or either of those options with an additional no opinion option. Any of those alternative scales could have been analyzed as a categorical, rank-ordered, and/or continuous scale. But, the scale the 1 4 6 10 researcher chose to use is neither fish nor fowl, and must have been very difficult indeed to analyze and interpret.

For more on Likert scales see Brown (in press), or the following websites: [Expired Link] trochim.human.cornell.edu/kb/scallik.htm or [Expired Link] http://lomond.icbl.hw.ac.uk/ltdi/cookbook/info_likert_scale/. For an example of Likert-scale questionnaires used in a Japanese-language needs analysis at the University of Hawai'i at Manoa (and how results can be reported from them), see Iwai, Kondo, Lim, Ray, Shimizu, & Brown, 1999 at <http://www.lll.hawaii.edu/nflrc/NetWorks/NW13/>

References

- Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. London: Cambridge University Press.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall.
- Brown, J. D. (translated into Japanese by M. Wada). (1999). *Gendo kyoiku to tesutingu*. [Language teaching and testing]. Tokyo: Taishukan Shoten.
- Brown, J. D. (2000). *Using surveys in language programs*. Cambridge: Cambridge University Press.
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. Rowley, MA: Newbury House.
- Iwai, T., Kondo, K., Lim, D. S. J., Ray, G., Shimizu, H., & Brown, J. D. (1999). *Japanese Language Needs Assessment 1998-1999*. (NFLRC NetWork #13). [HTML document]. Honolulu: University of Hawaii, Second Language Teaching & Curriculum Center. Retrieved April 30, 1999. Available at www.lll.hawaii.edu/nflrc/NetWorks/NW13/[Expired Link]

HTML: http://www.jalt.org/test/bro_7.htm **PDF:** <http://www.jalt.org/test/PDF/Brown7.pdf>