Using Likert-Type Scales in the Social Sciences

James T. Croasmun
Lee Ostrom

Abstract

Likert scales are useful in social science and attitude research projects. The General Self-Efficacy Exam is a test used to determine whether factors in educational settings affect participant's learning self-efficacy. The original instrument had 10 efficacy items and used a 4-point Likert scale. The Cronbach's alphas for the original test ranged from 0.76 to 0.90. A 5-item Likert scale was created from this instrument by first adding a "3 = neutral/undecided" option and also by adding five negatively-worded items to the instrument. The instrument was piloted with 20 participants. The Cronbach's alpha for this pilot study was 0.87. The instrument was subsequently used in a large research study, and the Cronbach's alpha was found to be 0.88. This yielded an instrument that showed strong internal consistency.

Introduction

Rating scales are commonly used in the social sciences and with attitude scores. Such instruments often use a Likert-type scale. A Likert-type scale "requires an individual to respond to a series of statements by indicating whether he or she strongly agrees (SA), agrees (A), is undecided (U), disagrees (D), or strongly disagrees (SD). Each response is assigned a point value, and an individual's score is determined by adding the point values of all of the statements" (Gay, Mills, & Airasian, 2009, pp. 150-151). A Likert rating scale measurement can be a useful and reliable instrument for measuring self-efficacy (Maurer, 1998). This type of scale was developed by Rensis Likert (1931), who described and then developed this technique for the assessment of attitudes.

For this study, a modified Likert-type scale was used with the General Self-Efficacy Exam to measure if a certain teaching method could have an effect on the self-efficacy of adult learners in college science courses. This article describes how the Likert scale and the number of items for this existing instrument were modified for use in studies and how data were gathered to confirm the reliability of the modified instrument.

Likert-Type Scales

Likert scales provide a range of responses to a statement or series of statements. Usually, there are 5

James T. Croasmun is the Curriculum Development Director at Brigham Young University in Rexburg, Idaho. Lee Ostrom is a professor at the Idaho Falls Center for Higher Education at the University of Idaho, Idaho Falls, Idaho.

categories of response ranging from 5 = strongly agree to 1 = strongly disagree with a 3 = neutral type of response (Jamieson, 2004). However, there is a debate among researchers concerning the optimum number of choices in a Likert-type scale. There are some researchers who prefer scales with 7 items or with an even number of response items (Cohen, Manion, & Morrison, 2000). Symonds (1924) implied that the optimal reliability is with a 7-point scale. If there are more than that, the increases in reliability would be so small that is would not be worth the effort to analyze the difference or develop the instrument.

Much research has been conducted on the subject of Likert scale items or categories, and there have been many seemingly contradictory findings. For example, Guilford (1954) stated that the optimal number of categories is a matter of empirical determination depending upon the situation. Mattel and Jacoby (1971), however, determined that the reliability and validity of an instrument is not affected by the number of scale points used for the items. Ray (1980) countered Mattell's (1971, 1972) studies by questioning the adequacy of their sampling that used unmatched groups of students. Thus, if a sub-sample were particularly heterogeneous, the answer format being responded to might appear to have artificially low reliability. Ray (1980) also determined that there was a significant difference between the differently constructed Likert scales. Increasing the number of Likert items from 3 to 5 contributed to a higher internal reliability (1951) and extra discriminating power.

When using Likert-type scales, it is essential that the researcher calculates and reports Cronbach's alpha coefficient for internal consistency reliability. Internal consistency reliability refers to the extent to which items in an instrument are consistent among themselves and with the overall instrument; Cronbach's alpha estimates the internal consistency reliability of an instrument by determining how all items in the instrument relate to all other items and to the total instrument (Gay, Mills, & Airasian, 2006, pp. 141-142). The researcher should sum the scales for data analysis and should not worry about analyzing the individual items in the scale. "If one does otherwise, the reliability of the items is at best probably low and at worst

unknown. Cronbach's alpha does not provide reliability estimates for single items" (Gliem & Gliem, 2003).

Since they have no neutral point, even-numbered Likert scales force the respondent to commit to a certain position (Brown, 2006) even if the respondent may not have a definite opinion. Odd-numbered Likert scales provide an option for indecision or neutrality. By giving responders a neutral response option, they are not required to decide one way or the other on an issue; this may reduce the chance of response bias, which is the tendency to favor one response over others (Fernandez & Randall, 1991). Respondents do not feel forced to have an opinion if they do not have one.

Using a mid-point item has been shown to affect the data. Preliminary results should be considered in their context; when surveying a population to ascertain opinion, then the inclusion or omission of a mid-point can alter the results considerably. The debate continues, and the explicit use of a mid-point is largely one of individual researcher preference (Garland, 1991). The use of both positively- and negatively-worded items in survey instruments has also been advocated for many years (Nunnally, 1978; Spector 1992) to avoid response bias.

Negatively-worded items are added to the scale to act as "cognitive speed bumps that require respondents to engage in more controlled, as opposed to automatic, cognitive processing" (Chen, Dedrick, & Rendina, 2007). Using negatively worded questions to minimize response bias is based on the crucial assumption that the items worded in the opposite ways are measuring the same concept that the positively worded items are measuring (Chen et al., 2007). Barnette (2000) found that Cronbach's alpha was higher and accounted for at least 10%, and in one case 20%, higher internal consistency as compared with any of the three conditions in which negatively-worded stems were used.

Method

The General Self-Efficacy Exam (GSE) was altered for this study. These modifications were made based on the research that has been conducted on the subject. The original GSE is a self-reporting, confidential question-

naire that measures student self-efficacy. Participants normally would be asked to respond to 10 efficacy items in the GSE that are based on a 4-point Likert scale. The GSE has demonstrated internal consistency through Cronbach's alpha. Schwarzer (2002) reported results from samples in 23 nations in which Cronbach's alphas ranged from .76 to .90 with the majority in the high .80s.

The final version of the modified GSE that was used in this study is a 15-question survey that uses a 5-point Likert scale. Keeping the 10 questions already in the survey, 5 questions were randomly chosen to be worded negatively and to be then placed after every 2 positive questions. A mid-point option was added to the scale was so that the scale was as follows: 1 = Not at all true, 2 = Hardly true, 3 = Undecided/Neutral, 4 = Moderately true, and 5 = Exactly true; this labeling is consistent with established guidelines for using surveys (Alreck & Settle, 2003). To score the instrument, the values of the responses on the negative items were reversed so that the values were as follows: 5 = Not at all true, 4 = Hardly true, 3 = Undecided/Neutral, 2 = Moderately true, and 1 = Exactly true.

Data

This instrument was tested on a pilot group of 20 people. They were asked to fill out the 15-question, 5-point Likert scale survey. After analyzing their responses with an SPSS statistics program, the Cronbach's alpha was found to be .87, which suggested strong internal consistency. Four months later, the same instrument was used with 80 people in a pre-test and post-test research design. The Cronbach's alpha for this larger group was .88.

Discussion

The 15 items in the modified GSE were reliable and consistent and were able to be used with confidence in a research project that measured the self-efficacy of students in a lecture-based science class and a highly interactive science class. The ordering of the questions may have had an effect on the student's ratings, but the questions were not shuffled to determine if this were the case. According to Alreck and Settle (2003), it would not have been wise to put all the negatively-worded questions together nor to put the negatively-worded questions next to their positively-worded counterparts.

The survey used in this study was built upon previous work, but Trochim (2006) outlined a process for creating a Likert scale from scratch. First, define the focus. Likert scales are unidimensional, and it is important to focus on what exactly you are trying to measure. Next, generate a set of potential scale items and then have a set of judges rate the items. To further narrow down the items, he recommended throwing out items that have a low correlation to the total score across all items. One can also get the average rating for the bottom and top quarter of judges and then do a *t*-test on the difference between the two. Items with higher *t*-values are good discriminators and should be kept.

While this is a valid method for constructing survey items, there was a small window of time in which to select and use a survey. Therefore, the survey was built upon the 10 survey questions created by Schwarzer which have been used for over two decades with high reliability and validity (Leganger et al., 2000). This modified GSE survey was tested on the same kinds of people that were included in the main study with the intention of discovering unanticipated problems with the wording of the questions. Those who completed the survey seemed to understand the questions and gave useful answers.

Conclusion

Creating a Likert scale instrument that showed internal reliability was very rewarding. This modified instrument that was developed was a derivative of Schwarzer's popular self-efficacy scale, which has yielded high internal consistency. Building a survey from scratch could be done following the principles outlined by Trochim although it would take longer to do so rather than to use an established instrument. There are many resources available for those who wish to make a custom instrument for a particular research project. It is hoped that others will use this modified GSE freely in their research on self-efficacy.

References

Alreck, P., Settle, R. (2003). *The survey research handbook* (3rd ed.). New York: McGraw-Hill.

Barnette, J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement, 60*(3), 361-370.

Barnette, J. (2001). Likert survey primacy effect in the absence or presence of negatively-worded items. *Research in the Schools, 8*(1), 77-82.

Brown, J.D. (2000). What issues affect Likert-scale questionnaire formats? *JALT Testing & Evaluation SIG*, *4*, 27-30.

Chen, Y., Rendina, G., & Dedrick, F. (2007). *Detecting effects of positively and negatively worded items on a self-concept scale for third and sixth grade elementary students*. Online Submission, Paper presented at the Annual Meeting of the Florida Educational Research Association (52nd, Tampa, FL, Nov 14-16, 2007). Retrieved Oct 6, 2010 from http://www.eric.ed.gov/ERICWebPortal/content delivery/servlet/ERICServlet?accno=ED503122

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London: Routledge Falmer.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

Fernandes, M., & Randall, D. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics, 10* (11), 805-807.

Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin*, 2, 66-70.

Gay, L. R., Mills, G. E., & Airasian, P. (2009). *Educational research: Competencies for analysis and applications*. Columbus, OH: Merrill.

Gliem, J., & Gliem, R. (2003). Calculating, interpreting, and reporting Cronbach's Alpha Reliability Coefficient for Likert-type scales. *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education.* Retrieved October 6, 2010, from http://hdl.handle. net/1805/ 344 .

Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education, 38*(12), 1217-1218.

Leganger, A., Kraft, P., & Røysamb, E. (2000). Perceived self-efficacy in health behaviour research: Conceptualisation, measurement and correlates. *Psychology & Health*, *15*(1), 51-69.

Likert, R. (1931). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 1-55.

Maurer, T. J., & Pierce, H. R. (1998). A comparison of Likert scale and traditional measures of self-efficacy. *Journal of Applied Psychology, 83*(2), 324-329.

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement, 31*(3), 657-674.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ray, J. (1980). How many answer categories should attitude and personality scales use? *South African Journal of Psychology, 10*, 53–4.

Schwarzer, R. (2002). The General Self-efficacy Scale (GSE). Department of Health. *Psychology*. Freie Universität, Berlin.

Spector, P. (1992). *Summated rating scale construction: An introduction.* Newbury Park, CA: Sage.

Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. *Journal of Experimental Psychology*, *7*, 456-461.