Expert Systems with Applications 40 (2013) 1609–1618

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A novel Likert scale based on fuzzy sets theory

Qing Li

Mechanical Engineering Department, University of New Haven, 300 Boston Post Road, West Haven, CT 06516, USA

ARTICLE INFO

Keywords:
Likert scale
Fuzzy sets theory
Fuzzy Likert scale
Information distortion
Information lost
Consensus

ABSTRACT

The Likert method is commonly used as a standard psychometric scale to measure responses. This measurement scale has a procedure that facilitates survey construction and administration, and data coding and analysis. However, there are some drawbacks in the Likert scaling. This paper addresses the information distortion and information lost arising from the closed-form scaling and the ordinal nature of this measurement method. To overcome these problems, a novel fuzzy Likert scale developed based on the fuzzy sets theory has been proposed. The major contribution of the fuzzy Likert approach is that it permits partial agreement of a scale point. By incorporating this capability into the measurement process, the new scale can capture the lost information and regulate the distorted information. A quantitative analysis based on the concept *Consensus* has proven that the new scale can provide a more accurate measurement. The implementation feasibility and the improved measurement performance of the fuzzy Likert scale have been demonstrated via a simulation study on a low birth weight analysis.

1. Introduction

Likert scaling, originally introduced by Rensis Likert in 1932, is the most widely used psychometric scale in survey research. It asks respondents to indicate their levels of agreement with a declarative statement. For a 5-point Likert scale, for example, each scale point could be labeled according to its agreement level: 1 = strongly disagree (SD), 2 = disagree (D), 3 = neither disagree nor agree (NN), 4 = agree (A), and 5 = strongly agree (SA). Depending on what is being measured, the scale labels may be worded differently. When measuring frequency, for instance, labels like "never-always" can be used; when measuring attitude, belief, or characteristic of the respondent, labels like "not very much-very much" are suitable. A well designed Likert scale should state the opinion, attitude, or belief being measured in clear terms and use the appropriate wording for scale points.

Likert scales have been widely used to measure observable attributes in various social science measurement areas. Examples of measured variables include fondness of music education (Orr & Ohlsson, 2005), organizational behavior in learning organization (Kiedrowski, 2006), satisfaction of journal quality in library science (Yue, Wilson, & Boller, 2007), effectiveness of drugs in pharmaceutics (Buncher & Tsay, 2006), patient advocacy in hospital (Seal, 2007), routine prioritization in dentistry (Postma, 2007), customer attitudes towards labeling in nutrition (Lindhorst, Corby, Roberts, & Zeiler, 2007), and athlete characteristics in sports (Brown, Guskiewicz, & Bleiberg, 2007), to name a few recent studies in the endless application list. In addition, Likert scales have also been applied to measure latent constructs that are not directly observable. For example, Springer (1998) used a Likert scale to examine adolescent concerns that foster runaway behavior, Copeland (2003) explored the problems faced by young women living in disadvantaged conditions, Acharya, Lee, and Im (2006) identified conflicting factors in construction projects, Bañuelas and Antony (2007) developed a stochastic analytic hierarchy process in operational research, and Li, McCoach, Swaminathan, and Tang (2008) applied a Likert scale to develop an instrument to measure student perspectives of engineering education.

The popularity of the Likert method comes from a number of facts. First, a Likert scale can be easily constructed and modified. Second, the numerical measurement results can be directly used for statistical inference. Last but not least, measurements based on Likert scaling have demonstrated a good reliability. In general, with Likert scaling researchers can collect and analyze a large quantity of data with less time and effort. Despite these advantages, however, Likert scales have several weaknesses.

One of the major problems comes from the debate about whether a Likert scale is ordinal or interval (Jamieson, 2004). Although Rensis Likert himself assumed that the Likert method has an interval scale quality, many consider Likert scaling as ordinal in nature (Hodge & Gillespie, 2003; Pett, 1997). A conventional interval scale implies that the differences between any two consecutive scales reflect equal differences in the variable measured. However, as Cohen, Manion, and Morrison (2000) argued, it is illegitimate to assume that the intensity of feeling between "strongly disagree" and "disagree" is equivalent to the intensity of feeling between other consecutive categories on a Likert scale. In fact, it is problematic to treat the Likert method either way. As Russell

E-mail address: cli@newhaven.edu

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.09.015

and Bobko (1992) put it, "Likert scales fail to approximate intervals of ordinal data". A typical ordinal scale can measure the orders of the responses but tells nothing about the intervals between responses. On the other hand, a typical interval scale forces an equal interval between consecutive scales. The former scale leads to information lost during measurement, while the later results in information distortion.

The other weakness of the Likert scaling comes from its closed response format (Hodge & Gillespie, 2003; Orvik, 1972). The respondents are forced to make a choice from the given options that may not match their exact responses. They have to either select an answer from an insufficient range of responses or respond to an "acceptable" answer in the closed format. This miss-matching further worsens the information distortion problem.

In summary, a significant amount of information is lost and/or distorted due to the built-in limitations of the Likert method. Over the years many researchers have tried to solve these problems. To address the information lost problem, Chang (1994) recommended increasing the scale points on a Likert scale. (Russell & Bobko, 1992) also suggested that the more scale points, the closer a Likert scale can approximate a continuous measure, and thus more information can be captured. However, criticism comments that respondents may have difficulty to accurately resolve their intensity of feelings into many scale categories. Through a comparison of 4- and 6-point Likert scales, Chang (1994) found that more scale points may actually increase the measurement error because respondents can be confused by too many response categories. In addition, Chang (1994) also pointed out that a longer response option list may intensify "laziness" in responding questionnaire. Therefore, this modification method can increase several typical "primacy effects" in Likert scaling, such as the response-order effect (increasing the tendency for respondents to select the first response available to them on the answer scale, Chan, 1991), donkey vote effect (selecting the same response for all questions, Ray, 1990), or central tendency effect (choosing the neutral response, Brown, 2000).

Later on Albaum (1997) proposed a two-stage Likert scale as an alternative to the traditional scale. In this alternative scale, the first stage measures the agreement (agree/disagree) to a statement. The second stage measures the intensity of agreement (strong or weak) to the statement. It appears that the two-stage Likert scale can capture more extreme positions than the traditional Likert scale. In a sense, this design is effective at reducing the central tendency effect. However, it is not clear how this method can collect more information between the extreme positions than the traditional method.

More recently Hodge and Gillespie (2003) proposed a "phrase completion" Likert scale approach that uses a sentence completion format to measure agreement. For example, the statement "My religious beliefs affect every aspect of my life" is replaced with an incomplete sentence fragment "My religious beliefs affect" plus two phrases "No aspect of my life" and "Absolutely every aspect of my life" that are linked to a scale ranging from 0 to 10. These two phrases, which complete the sentence fragment, anchor each end of the scale. This phrase completion method exhibits some advantages to the traditional Likert scale. By specifying the underlying theoretical continuum in the response, this method can capture more detailed information. It is also proven that the reliability and the inter-item correlations of the measurement are higher. However, a long list of response choices can lead to the typical "laziness" phenomenon (Chang, 1994). Furthermore, as the authors commented, the two-phrase completion format may also confuse respondents and present an operational challenge.

Good efforts have been spent on the development of alternatives to Likert scales for better measurement results. However, these alternatives have not circumvented the built-in limitations. In most practices in the current social science measurement society, these limitations are simply ignored. In this paper, a novel Liker scale designed based on fuzzy sets theory has been proposed. The fuzzy sets theory offers psychometricians a new interpretive algebra, "a language that is half-verbal-conceptual and half-mathematical-analytical" (Ragin, 2000). This interpretive mathematical language can transform a discrete ordinal variable into a continuous variable while retaining the semantic meaning. It thus provides us an idea tool to capture the interval details of ordinal variables in an open response format. With that capability, it is possible to reduce information lost and decrease information distortion during measurement. A consensus model originated from communication theory (Tastle & Wierman, 2007) has been applied to rigorously prove that the proposed fuzzy Likert scale can provide a more accurate measurement than the traditional Likert method. A logistic regression simulation study based on a low birth weight analysis has demonstrated the implementability and effectiveness of the novel fuzzy Likert scale.

2. Introduction to fuzzy sets theory

The concept of *fuzzy sets* was first conceived by Lotfi Zadeh, an engineering professor at the University of California at Berkley, to deal with reasoning that is approximate rather than precise (Zadeh, 1965). Since then, fuzzy sets have been used in many engineering fields to address a variety of problems, both mundane and abstract (Zimmermann, 1996). The ever expanding applications of fuzzy sets have ranged from expert systems (Zimmermann, 1987), manufacturing systems (Gien, Jacqmart, Seklouli, & Barad, 2003), operational research (Zimmermann, 1983), to stock market (Zopounidis, Pardalos, & Baourakis, 2001). Most of the literature in fuzzy sets applications is concerned with the development of smart machines that can act automatically in the face of ambiguity or complexity (Jamshidi, Titli, Zadeh, & Boverie, 1997).

Although fuzzy sets have found a great success in engineering, their impact in social sciences has been rather limited. Several scholars have attempted to introduce fuzzy sets concepts to social science community (Ragin, 2000; Smithson, 1987; Smithson & Verkuilen, 2006). The majority has yet to recognize the potential of fuzzy sets for transforming social science methodology. By incorporating fuzzy sets into the traditional forms of qualitative and quantitative analysis in social sciences, we are equipped with a powerful mathematical model that is able to retain the substantial meaning of the underlying latent constructs without losing analytic rigor. This capability of fuzzy sets can bring the measurement methodology in social sciences up to a whole new level.

2.1. Fuzzy sets

Since our language abounds in imprecise and fuzzy information by nature, fuzzy concepts and fuzzy reasoning are commonly seen in social science research. For example, a simple statement like "This family income is high" can be fuzzy because the linguistic label "High" is not precisely defined. Different people will have different standards in labeling "High". However, it is difficult to translate this statement into a more precise language without losing some of the semantic meaning. For instance, an alternative to this statement could be "This family income is $100K/year". It does indicate an exact number but no longer explicitly delivers the semantic meaning "The income is 'high'".

How about we include both exact number and semantic meaning into the statement: "If a family has an income of $100K/year, this family income is a high"? This seems to be crystal clear now. But what if a family earns $90K/year, does it still belong to the "High family income" category? The question becomes fuzzy again.