Chapter One

Introduction

‘Measurement’ is defined as assignment of numeric values to objects or events according to rules. Relationships among such assigned values reflect analogous relations among objects. Rating scales and Likert scales are the most popular psychological measurement schemes that depend on human judgment (Guilford, 1954; Nunnally & Bernstein, 1994; Zimmermann, 1996). Likert scales, using the standard set of response options representing the degree of agreement instead of descriptive terms, do not differ fundamentally from rating scales. Both scaling schemes assume that the human observer is good at quantitative observation, and assign of numbers or objects to reflect degrees of traits or statements being measured (Guilford, 1954; Hopkins, Stanley, & Hopkins, 1990).

The merits of applying rating scales include conciseness, time efficiency, breadth of applications and minimum training of raters. However, this approach has been criticized on the grounds that it is too simplistic. First, raw scores of a rating scale fail to meaningfully measure objects. Meaningful measurement must be linear, enabling arithmetic and linear statistics to be applied. Unfortunately, raw rating scale scores are merely ordinal, non-linear and sample dependency (Wright, 1999). Second, the options used in the rating scales, without clear and mutually exclusive distinctions, could be viewed as “linguistic variables”. Linguistic variables are variables whose values are not numbers but words or sentences in a natural or artificial language (Zimmermann, 1996). Lacking clear definitions for the variables, the arithmetic performed on linguistic variables is beyond the capability of traditional binary crisp logic. Therefore, the newly developed fuzzy logic is the preferred solution to measurement.

Concerning the issue of meaningful measurement, raw scores suffice to accomplish a meaningful measurement. Summarizing from Thorndike’s “new" measurement in the 1920s, Campbell's “fundamental measurement”, Guttman's “conjoint transitivity”, Rasch's “specific objectivity”, and Wright’s “objective measurement”, psychological measurement must be unidimensional, linear, invariant and objective (Rasch, 1960; Wright, 1997; Wright & Linacre, 1989; Wang, 1996). Unidimensionality, the universal characteristic of all measurement, means that measurement of any object or entity describes only one attribute of the object measured (Thurstone, 1931). Linearity refers to equality of measurement unit, that is, measurement must be linear allowing arithmetic operations to be performed on them. Invariance means that the measurement maintains its properties when the unit of measurement is changed. Objectivity means sample-free and test-free, that is, item calibrations must not depend on the respondent are used to estimate them, and measurements must not depend on which items are taken.

However, raw rating scale scores are non-linear, and sample and test dependent. The non-linearity of the raw scores, in favor of central scores and against extreme scores implies that applying any linear statistical method, such as general linear model (GLM), Pearson's product moment correlation, factor analysis or structural equation modeling (SEM), produces systematically distorted results (Wright & Masters 1982). To attain objective, meaningful measurement, raw scores must be transformed into linear measures to enable subsequent analysis and inference.

The Rasch model, a simple logistic latent trait model for analyzing test data initiated by Danish mathematician Georg Rasch in 1960, improves measurement in the social and behavioral sciences. The Rasch model specifies the probability of a correct response to an item as an exponential function of the difference between person ability and item difficulty (Masters, 1982; Andrich, 1988). Furthermore, since the raw score is the sufficient statistics for the personal ability parameter in Rasch model, 'objective comparisons' are possible since comparison of two person parameters does not depend on the details of the items used to make that comparison, and the comparison of two item parameters does not depend on the details of the persons taking those items (Masters, 1982). Being unidimensional, linear, invariant, and objective, the Rasch model makes fundamental measurement and objective comparison in psychology possible.

Rasch's model is being extended to address every conceivable raw observation: dichotomies, rating scales, partial credits, binomial and Poisson counts (Masters & Wright, 1984). The partial credit model (PCM, Masters, 1982), a unidimensional Rasch model for polychotomous items, parameterizes the difficulties of a series of categories, called "steps", in each item. A distinguished feature of PCM is that item steps may vary between items. The PCM is a member of the family of latent trait models with parameter separability, permitting "specifically objective" comparisons of person and items (Masters, 1982; Masters & Wright, 1984).

Regarding the second debate, the alternatives presented in the options alongside the rating scales are not numerical variables. These alternatives are linguistic variables whose values are not precise numbers but words in human languages. For instance, "speed" is a linguistic variable if it takes a value such as slow, moderate or fast. Linguistic variables can approximately characterize phenomena that are too complex or too ill-defined to be describable in conventional quantitative terms. Linguistic variables are less specific but more comprehensive than numerical ones in daily life. For instance, people may say, “It’s awfully hot today,” a straightforward imprecise linguistic term, rather than, “The temperature is 35 degrees Centigrade,” a precise numerical statement, to comment about the weather. Similarly, a question about sadness adapted in the Beck Depression Scale II (BDI-II, Beck, Steer, & Brown, 1996) such as “I did not feel sad,” ” I felt sad much of the time,” ” I felt sad all the time,” and “I’m so sad or unhappy that I can’t stand it,” are linguistic variables since these terms are not clearly defined and no definite boundaries exist between, for example, “much of the time” and “all the time”.

Furthermore, the distinctions between two adjacent alternatives may be so polarized or extreme that choosing one alternative to represent one’s state is burdensome. Considering the example quoted above, the discrepancy between two adjacent alternatives such as “I did not feel sad,” and “I felt sad much of the time” seems so strong that examinees who only felt sad occasionally would not be easily able to select an alternative. Under such circumstance, assuming someone entirely belongs to one particular alternative may be debatable. Such an assumption is a crisp set view originated in Aristotle’s binary logic, where each individual can be dichotomized into member of a set (those who certainly belong to the set) and nonmember (those who certainly do not). Inheriting from this view, test constructors ask examinees to choose one alternative (set) in each item. However, people generally feel depressed, satisfied, happy or comfortable in the continuum within two opposing extremes rather than a yes-or-no dichotomy. Therefore, human thinking is multi-valued, transitional and analogue, instead of bi-valued, clear-cut, and digital. Fuzzy set theory, providing a systematic framework for dealing with the vagueness and imprecision in human thoughts, is a powerful tool to analyze and animate human thinking (Dubois, Ostasiewicz, & Prade, 2000). Nevertheless, few fuzzy set theory applications have been found in psychometric studies.

Fuzzy set theory, initiated by Lotfi Zadeh in 1965, initiated the grand scientific paradigm shift from the traditional crisp-set view to the alternative fuzzy-set view in the twentieth century (Klir &Yuan, 1995). According to the traditional view, based on Aristotelian binary logic, science should strive for certainty, by being crisp, precise, and deterministic, and uncertainty is undesirable and unscientific. On the contrary, the alternative modern fuzzy-set view perceives uncertainty essential and valuable to science (Klir &Yuan, 1995; Kosko, 1993; Zimmermann, 1996).

Considering the traditional crisp-set view, classical logic is crisp, precise, and certainty in character. ‘Crisp’ herein means dichotomous, that is, assuming a sharp, unambiguous boundary between membership and nonmembership of a given set, each individual can be dichotomized into member (those that certainly belong in the set) and nonmember (those that certainly do not). Precision implies that the model is unequivocal with no ambiguities, and that each measurement is accurate to a given numerical precision. Certainty indicates that a model's structures and parameters are clearly known, and that a proposition either clearly holds or does not (Dubois et al., 2000; Zimmermann, 1996).

However, the real world is often multi-valued, vague and uncertain. The philosopher Russell pointed out that traditional logic employing precision symbols is not applicable to this terrestrial life but only to an imaginary celestial existence (Kosko, 1993). Real life situations, particularly in mental representations and natural languages, tend to be characterized by transitions rather than yes-or-no dichotomies. Concepts represented in human languages or mental representations make more sense than precise symbols but are very vague, without well-defined boundaries for the object classes to which they. Classical logic cannot account for concepts described in natural language, such as cold weather, young men and expensive cars. On the contrary, the capacity of fuzzy sets to express gradual transitions from membership to non-membership is very useful. For instance, any cloud cover of 20% or less might be considered as sunny. Assuming a crisp-set view, a cloud cover of 21% is not sunny since the criterion dichotomizes each weather condition as "sunny" or "not sunny". This rough dichotomy is unrealistic since 1% of cloud cover hardly seems like a distinguishing characteristic between sunny and not sunny. In the fuzzy-set view, the term "sunny" may introduce vagueness by allowing a gradual transition from "sunny" and "not sunny" (Klir &Yuan, 1995).

Considering the fuzzy set theory, the degree to which an element belongs to a given set, denoted by "membership", is a continuous value, gradually changing from zero to one. Therefore, a fuzzy set can be defined mathematically by assigning a value representing its membership grades to each possible individual in the universal discourse. The membership function, or the character function of a fuzzy set, corresponds to the level of similarity, likelihood, or compatibility with the concept represented by the fuzzy set. These memberships are very often denoted by real-number values ranging between 0 and 1. The traditional dual logic corresponds to the special cases in a fuzzy set in which the membership degree is either of the extreme values, 0 or 1 (Bilgic & Tuksen, 2000; Dubois et al., 2000; Zimmermann, 1996).

Fuzzy sets theory and its derivative, possibility theory, have advanced in many disciplines. For instance, the theory has been applied in artificial intelligence, computer science, decision theory, logic and pattern recognition (Dubois et al., 2000). Since fuzzy sets theory provides a systematic framework for dealing with the vagueness and imprecision inherent in the human thought process, it should be beneficial in psychometric investigations. First, variables of interest in psychology are poorly-defined, latent, and imprecise, corresponding to the vagueness in fuzzy set theory. Second, each item presented in psychological inventories could be regarded as a linguistic variable. However, in contrast with the many engineering studies discussing fuzzy set theory, only a few such works have been published in psychological measurement. These works include a series of studies conducted by Berlin Wu and his associates (Wu, 1995; Wu & Lin, 2002a; 2002b), which revealed that the fuzzy set approach is more efficient in predicting human behavior and public opinions than traditional logic. Furthermore, another series of studies by Yuan-hong Lin(Lin , 2001; 2003a; 2004) demonstrated that, based on computer simulations and real a case study, the fuzzy set approach is more reliable and accurate than the traditional scoring of Likert scale.

As well as administration in the scaling approach, the fuzzy set theory is also utilized in statistical analysis. Fuzzy statistical analysis is a powerful data analysis tool. Applications of fuzzy statistical analyses have recently been extended to traditional statistical inferences and methods in fields such as education, psychology, economics and public administration. Wu & Hsu (2004) developed a fuzzy time series model to overcome the bias of stock market systems. Wu and Yang (1998) demonstrated the concepts of fuzzy statistics and applied them to social surveys; Wu and Tseng (2002) employed the fuzzy regression approach estimation to analyze Taiwan's monitoring index of economics. Yang, Hwang, & Chen (2004) modified the conventional similarity measure for mixed data in fuzzy clustering to produce better classification results in Taiwan's automobile market. Hung & Yang (2005) applied fuzzy clustering to Taiwanese tea quality evaluation.

Statement of Problem

The previous paragraph suggests that criticism of the traditional scaling method mainly highlights the measurement issue and debates about vagueness of variable definitions. A more accurate scaling method which integrates the PCM (the polychotomous Rasch model that transforms raw scores into objective measurement), with the fuzzy set theory (the mathematical framework in which vague conceptual phenomena can be delicately studied), seems to be feasible, but so far has not been proven empirically.

Therefore, this study proposed a new scaling method, fuzzy partial credit scaling (FPCS). The alternatives for each scaling item are considered as fuzzy numbers. In contrast with the traditional crisp set view where an examinee belongs to exactly one alternative (set), FPCS enables an examinee to belong to many alternatives (sets) conjointly. Percentages assigned by examinees represent the membership degrees to which they belongs to an alternative (set). The membership degrees may express the grade of similarity, likelihood, utility or compatibility with the concerning set and, in turn, allows effective statistical analysis for fuzzy data.

Instead of successive integrals in traditional rating scale scoring, FPCS utilizes “step parameters” estimated from the PCM. First, triangular fuzzy numbers are constructed using step parameters to characterize distributions of each alternative value, “fuzzifying” the crisp data to reveal uncertainty. Next, we adopt center of gravity (COG) method to “de-fuzzify” the fuzzy number into a scalar to denote the fuzzy number value. Then, the “observed fuzzy scores” defined in FPCS is computed as the sums of fuzzy number values weighted by membership degrees.

The instruments in this study was the Beck Depression Scale II (BDI, Beck, 1996), the most widely used and investigated self-report measure of depression for clinic samples, The total sample in this study comprised two different populations: a depressed sample, recruited from psychiatric outpatients who were diagnosed depression, and a non-depressed sample recruited from college students. The depressive symptoms were categorized into serve, moderate, mild, partial remission and full remission according to the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, American Psychiatric Association, 1994) by psychiatric doctors.

The purpose of this study is to establish and validate the proposed scaling method, FPCS. To achieve this goal, a series of studies was conducted. To validate the FPCS, the differences between FPCS and traditional scaling approach were compared and discussed.

Study One. This study focused on the reliability of FPCS. Since this study proposes to modify the traditional scaling method, the superiority of the proposed scaling method must be assured. Concerning the reliability, Cronbach alpha coefficient and the squared correlation between the latent and the observed variables estimated via SEM were applied to investigate whether FPCS exhibit higher reliability than raw scores does.

Study Two. This study focused on the validity of FPCS. Since the validity of an instrument can be view as the accuracy of specified inferences made from its scores. Predictive validity was adopted to investigate whether validity of FPCS was higher than that of the raw score.

In this study, logistic regression and discrimination analysis were adopted to evaluate predictive validity. Logistic regression was applied to predict the odds of suffer depression, while discrimination analysis was used discriminate non-depression, depression with remission, and depression without remission.

Study Three. The main goal of this study was to discover whether fuzzy clustering is more desirable than traditional crisp clustering method. FCM, a possibility-based fuzzy statistical method was applied to the collected data to investigate whether FCM is a more accurate clustering approach than Wald's method and the *k*-means method.

Hypotheses

The following predictions have been made for this study:

Study One. Concerning the reliability, the analysis of internal consistency, Cronbach alpha coefficient, and confirmatory factor analysis (CFA) model of the depression scales of FPCS and raw score will be constructed and compared to each other. The following predictions were made:

1. Concerning the CFA approach, FPCS will exhibit higher reliability than the raw score, i.e., the proportion of variance in the variables accounted by the factor in FPCS will be larger than the raw score.
2. Concerning the internal consistency approach, the Cronbach α coefficient of FPCS will be higher then that of raw score.

Study Two. Logistic regression analysis and discrimination analysis were used to evaluate the predictive validity.

1. Concerning logistic regression, FPCS will yield more accurate predictions for depression than the raw score do.
2. Concerning discrimination analysis, more cases will be correctly classified by FPCS than by raw score.

Study Three. Fuzzy-based and crisp-based clustering analyses were applied to partition cases. Cluster validity indices were used to determine the optimal cluster number and exponential weight in FCM. The hypothesis made for this study is that fuzzy-based FCM will yield higher associations between original and classified groups than crisp-based approaches.