Chapter Five

Discussion

This study attempts to verify that the proposed scaling method, FPCS, is a more reliable, valid and accurate scoring schema than raw scoring. Overall, FPCS is consistently shown to be superior to raw score in terms of reliability, validity and clustering accuracy. The legitimacy of applying fuzzy set theory to psychological research was empirically demonstrated. The overall conclusion, discussion of results, limitations, and future directions are presented below.

I. Overall Conclusions

This study applied the Rasch model which makes certain assumptions about data. Rash analysis on BDI shows that five items in the original 21-item BDI did not demonstrate an acceptable fit and were eliminated. The analyses of Study One to Study Three were based on the 16-item BDI concerning the item-fit criterion.

The analytical results of Study One reveal that FPCS exhibited higher reliability of BDI than did raw scoring. That is, compared with raw score, scoring via FPCS produced a smaller measurement error leading to greater variance in BDI items explained by depression.

Analytical results of Study Two demonstrate that the predictive validity of FPCS was higher than that of the raw score. Logistic regression was employed to predict the odds of suffering depression, while discrimination analysis was used to distinguish among non-depression, depression with remission and depression without remission. These two statistical methods demonstrate consistent results that FPCS yielded higher predictive validity.

In Study Three, a possibility-based fuzzy statistical method, FCM, 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, both of which are crisp clustering approaches. Compared with the crisp-based clustering approach, fuzzy-based FCM yielded stronger associations between the original and classified groups. That is, FCM most accurately revealed the information inherent in the latent data structure.

In conclusion, FPCS is a more reliable, valid and accurate scoring schema than raw scoring. Restated, the FPCS scoring schema, combining fuzzy set theory with the Rasch model, is a promising scaling scheme for psychological measurement.

II. Results of Study One

It is almost impossible to deal with the issue of reliability without addressing the concept of true score. In classical test theory, an observed score on a measurement is the sum of true score and measurement error, and can be expressed symbolically as

$$X = T + E.$$

Where X represents the observed score; T represents the true score, and E represents the measurement error. The reliability coefficient is defined the ratio of true variance to total variance. The true score of an individual is considered as a personal parameter that remains constant, and can generally be regarded as the average score obtained over repeated measurements. Measurement error causes scores to vary over the tests. However, the measurement errors are never entirely eliminated, so they should be minimized (Nunnally & Bernstein, 1994).

Furthermore, a scoring schema of reliability should remove most measurement errors, thus increasing the ratio of true variance to total variance. In this section, the reliability coefficient was obtained by the SEM approach, in which the reliability was estimated from the proportions of variances in an observed variable that is accounted by the latent construct. The analytical results reveal that for each item (observed variable) in BDI, the variances explained by the latent construct depression computed using FPCS were higher than those computed via raw score. Additionally, the Cronbach α coefficient also shows that the internal consistency of FPCS is higher than that of raw score. In conclusion, therefore, the proposed scoring schema, FPCS, exhibited a higher reliability, eliminating more measurement errors, than the raw score.

The increased number of eliminated measurement errors is likely to be the result of the interval measure and fuzziness stemming from FPCS.

FPCS originates from PCM, a member of Rasch family, which is unidimensional, linear, invariant, and objective (Wright, 1999). FPCS applies PCM to generate “step parameters” from the difficulties of a series of categories in each item. The step parameters are then employed to build fuzzy triangle numbers to illustrate the distributions of alternatives presented in rating scales. Since PCM achieves an interval measure, and the calculation procedures of FPCS could be regarded as linear combinations of parameters in PCM weighted by membership degrees. FPCS is derived from PCM, which achieves an interval measure, therefore, FPCS is more linearity approximated in contrast to raw score.

A multi-valued fuzzy logic reduced measurement error system was introduced to achieve fuzziness. Measurement errors comprise systematic and non- systematic error components. Some systematic error components, such as the error of leniency and error of severity, could be eradicated by multi-valued fuzzy logic. The error of leniency indicates that the rater tend to rate those higher than they should, whereas the error of severity indicates that the rater tend to rate those lower than they should. These errors were presumably a constant tendency regardless of trait (Guilford, 1954).

Taking the item "Sadness" presented in BDI as an example, the four alternatives presented were "I do not feel sad", "I feel sad much of the time", "I am sad all the time", and "I am so sad or unhappy that I cannot stand it". Someone who only feels sad occasionally must choose between "I do not feel sad" (the first alternative) and "I feel sad much of the time" (the second alternative). However, the distinction between two adjacent alternatives is so polarized or extreme that selecting one alternative is burdensome. When a respondent is asked to choose precisely one alternative presented in the rating scales to describe his mood state or attitude, some force fitting and rounding off are inevitable. Such reduction might involve the error of leniency and severity, reducing the reliability of the instrument. This example demonstrates that traditional crisp logic was insufficient to reflect a person's state. Furthermore, the reliability of an instrument is diminished due to some force fitting or rounding off when a respondent is required to select precisely one category in a rating scale to denote one's state, attitude or opinions.

Summary

The analytical results of this section demonstrate that FPCS exhibited higher reliability of BDI than raw score. That is, scoring by FPCS generated a lower measurement error than raw scoring. Therefore, the increase in the variance in an item of BDI was explained by depression, the latent construct. Minimizing measurement error is essential to any measurement. In this respect, FPCS is a better scoring system than raw scoring.

III. Results of Study Two

Validity is an integrated evaluative judgment of the extent to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inference based on test score. That is, validity is an inductive summary of both the

existing evidence for, and the potential consequence of, the score interpretation and use. Hence, the inferences derived from the test scores are validated, rather than the instrument itself (Messick, 1989).

Psychological and educational measurements are inferred from scores, the coding schema of a measurement. Therefore, a valid scoring system should be able to calibrate the trait being measured to a scalar measure using items present in the scale. That is, a good scoring system should bridge the gap between the items and the trait, and reflect the magnitude of the trait correctly. Given these considerations, this study showed that FPCS is a valid scoring schema.

The analytical results of this study reveal that FPCS demonstrated higher validity than raw scoring provided an empirical support for that FPCS could precisely reflect human thinking. Psychiatric doctors diagnose mental disorders based on DSM-IV criteria, and BDI was built to comply with DSM-IV criteria to evaluate of the severity of depression. It is supposed that diagnoses performed by psychiatric doctors and scoring results from BDI were found to agree well, thus the validity of the scoring was established. In this respect, FPCS is a better scoring system than raw scoring.

The analytical results reveal that FPCS is a more accurate scoring method than raw scoring, probably because of its application of fuzzy logic. Since fuzzy logic was developed to handle the vagueness in human thinking, it can convey human thinking more accurately than can crisp logic. The following paragraph discusses the uncertainty inherent in human thinking and how crisp logic fails to account for these phenomena.

Information can involve many uncertainties, both numeric and non-numeric. The major forms of uncertainty are those arising from randomness, vagueness and imprecision. Probability theory and fuzzy logic are principal methodologies for dealing with uncertainty. Probability theory models uncertainty caused by randomness, also called variability. Nevertheless, probability theory rooted in binary logic is only concerned with precise binary outcomes, and fails to consider imprecision in human thought. Conversely, fuzzy logic is mainly concerned with uncertainty arising from vagueness and imprecision (Ross, Sellers, & Booker, 2002).

Uncertainty in psychological measurement involve not only randomness, but also vagueness and imprecision. Linguistic terms in natural languages have been utilized in psychological measurements to elicit human thinking. A psychological measurement is based on the assumption that the human observer can make precise quantitative observations (Guliford, 1954). Therefore, numbers or objects are often assigned to statements to reflect the degrees of traits or statements being measured. However, this assumption has been criticized because human thinking, like the linguistic terms which represent it, is not precise. Conversely, linguistic terms in

natural language are vague and imprecise (Dubois et al., 2000; Klir & Yuan, 1995;; Wu, 1995; Wu & Lin, 2002a).

Natural languages are abundant in vagueness. A proposition is vague when its meanings not fixed by a sharp boundary. That is, the possible statement of the proposition is not clear defined with respect to its inclusion (Ross et al., 2002; Wu & Lin, 2002a). Consider, for instance, the proposition, "I am sad", an item applied in BDI. Since the state of sadness gradually ranges between the two extremes, rather than yes-or-no dichotomies, it cannot be dichotomized into "sad" and "non-sad". Most linguistic terms applied in psychological measurement are vaguely defined like the foregoing example. Classical binary logic does not hold under these circumstances. By contrast, uncertainty due to vagueness adopts fuzzy set theory, which was developed to manage the vagueness inherent in natural language (Klir & Yuan, 1995). In fuzzy set theory, vagueness is described by degree of membership. Therefore, the vagueness of natural language can be expressed and analyzed using algorithms developed from fuzzy set theory.

Imprecision, which is also an attribute of natural language, means that any measurement yields results of limited accuracy. Imprecision can be associated with quantitative or qualitative data. An example of quantitative data might be the statement "I slept about 60 minutes less than usual". Also, an example for the qualitative data listed in BDI might be the statement "I slept somewhat less than usual".

Since information is random, vague, and imprecise, uncertainty is inevitable. Uncertainty signifies inability to determine whether a proposition holds. Modeling uncertainty refers to the quantification of partial beliefs regarding proposition (Dubois et al., 1997). Fuzzy set theory is widely recognized to be efficient for analyzing uncertainty, and has been successfully applied to artificial intelligence, computer science, decision theory, logic and pattern recognition (Dubois et al., 2000; Klir & Yuan, 1995;; Wu, 1995; Wu & Lin, 2002a). This study provided empirical evidence that fuzzy set theory is also beneficial) for analyzing psychological data, since it was found to yield high validity, the prerequisite for any psychological measurement.

Summary

Theoretically, fuzzy logic should handle vagueness and imprecision in human thinking better traditional crisp logic. Empirically, the validity investigation in this study proves the above theory. When vagueness in human thinking is inevitable, crisp (binary) logic, which describes real-world situations using a simplified mathematical model, trades accuracy for simplicity. When selecting one alternative among many

statements presented in rating scales to describe a person's mood state or attitude, some force fitting and rounding off are inevitable, resulting in some loss of information (Kosko, 1993). Therefore, the FPCS scheme, based on fuzzy logic, is a valid scoring schema for psychological measurement.

IV. Results of Study Three

The aim of this investigation was to study the applicability of fuzzy set theory to psychological measurement. In the previous experiments, scores of BDI were determined using fuzzy logic and were subjected to traditional probability-based statistical methods such as SEM, logistic regression and discrimination analysis. That is, the scores were "fuzzy" while the analytical techniques were "crisp". By contrast, in the following experiment, a possibility-based fuzzy statistical approach was conducted on the collected data.

FCM is an unsupervised and multi-membership technique. To compare FCM with other crisp clustering methods, the results of FCM were modified in two ways. First, the original sample membership was taken as the criteria to evaluate the classification accuracy. Second, each classified group was assigned to the group that acquired the highest membership degree. That is, the membership degrees were crispified binary values.

The primary aim of cluster analysis is to discover structure or information in data. With regard to psychological data, the results of clustering are intended to identify the structure inherent in latent psychological constructs. Traditional crisp clustering methods recognize and categorize patterns dichotomously. The probability-based clustering methods had shown impressive results in many aspects. Nevertheless, these methods have some limitations when the data structure was based on possibility rather than probability. Human thinking, traits, attitude and natural language applied to denote human perceptions are based on multiple-value instead of binary logic. Therefore, the membership of a particular group is a transition from 0 to 1 rather than a binary choice between 0 and 1. Considering the latent construct "depression" as an example, the depression state of a person is a transition from non-depressed to depressed. Therefore, whether a person belongs to the group "depression sufferers" is better represented by gradual membership rather than a yes-or-no dichotomy. This investigation justifies these theoretical inferences. Fuzzy-based FCM was found to yield stronger associations between original and classified groups than crisp-based clustering. Restated, FCM uncovers the information inherent in latent structure more

accurately than crisp-based clustering.

V. Limitations and Future Directions

The analytical results show the reliability, validity and classification accuracy of FPCS. However, several limitations should be noted.

A. IRT Models

First, the limitations inherent from IRT models must be considered. Although the various IRT models had advanced in fields of personality and attitude assessment, however, application of IRT in these domains is relatively rare. One explanation for the limited application of IRT models to psychological research may be that IRT models are based on strong assumptions, which are not be completely met by any of the data sets (Embretson & Reise, 2001; Hambleton & Swaminathan, 1985). Some controversy about IRT assumptions remain unsolved remain unsolved, particularly the assumption of unidimensionality. For instance, many of the statistical procedures proposed to assess dimensionality cannot distinguish between one- and two-dimension data sets (Hattie, 1985), as shown by the dimensionality assessment on BDI in this study. Additionally, many personality constructs do not refer to unidimensional variables but are multifaceted. In short, many personality assessment constructs are conceptualized on different levels (Embretson & Reise, 2001). Concerning the construct discussed in this study, depression is considered a multifaceted syndrome that manifests affective, cognitive, behavioral and vegetative symptoms. Additionally, the field of psychopathology has yet to reach a level of knowledge that would enable the best system of diagnosing and classifying for depression to be identified (Beckham, Leber, & Youll, 1995). Therefore, for future research, multi-dimensional IRT models (Wang, 2004) may be appropriate for analyzing depression.

B. Fuzzy Numbers and Defuzzification Methods

In this study, triangular fuzzy numbers were constructed using step parameters to characterize distributions of each alternative value. Future studies should try different shapes of fuzzy numbers, such as trapezoidal, L-R or Gaussian fuzzy numbers to discuss what fuzzy numbers could denote represent human thinking most accurately.

Furthermore, fuzzy number de-fuzzfication, also called fuzzy number

ranking, is still under discussion. Various ranking methods, including ranking using degree of optimality, α-cuts, centroid index and area measurement have been proposed (Chen & Huang, 1992). Apart from the center of gravity methods, other fuzzy number ranking schemes should be considered in future studies.