Kansei Engineering Vol.7 No.1 pp.81-86 (2007)

ORIGINAL ARTICLES

APPLYING FUZZY THEORY TO SCORING PSYCHOLOGICAL MEASUREMENTS

Sen-Chi Yu* and Yuan-horng Lin**

* Center for Teacher Education, Huafan University

** Department of Mathematics Education, National Taichung University

Abstract: The aim of this study was to propose and validate the new scaling method, fuzzy partial credit scaling (FPCS), which combines fuzzy set theory with the partial credit model (PCM) to score rating scales. To achieve this goal, the Chinese version of BDI (Beck Depression Inventory-II) was administrated to a depressed sample of patients and a non-depressed sample. The depressed sample consisted of 240 outpatients who were diagnosed as depressed by a psychiatric doctor, while 321 undergraduate students were recruited for the nondepressed sample.

In FPCS, triangular fuzzy numbers were generated by step parameters to characterize distributions of each alternative value. Next, the center of gravity (COG) method was applied to "de-fuzzify" the fuzzy number into a scalar. Then, the "observed fuzzy scores" defined in FPCS were calculated as the sums of fuzzy number values weighted by membership degrees for the following analysis.

The predictive validity issue of FPCS was investigated. Discrimination analysis was performed to classify the subjects according to the severity of depression into three categories: non-depression, depression with remission and depression without remission. The analytical results exhibited that, via FPCS, the probability of correct classification of severity of depression was raised from 71.2% to 80.7%. These two statistical analyses consistently show that FPCS exhibited higher predictive validity than did the raw score. That is, BDI scoring via FPCS makes more accuracy predictions for depression than raw score.

FPCS has been consistently shown to be superior to raw scoring in terms of reliability, validity, and clustering accuracy. This study has empirically shown that fuzzy set theory is applicable to psychological research.

Keywords: *Fuzzy partial credit scaling, Fuzzy set theory, Rasch model, Depression.*

1. INTRODUCTION

In classical test theory (CTT), "Method of successive integral", or "raw score", is most straightforward and popular scoring method in psychological measurements [1]. In method of successive integral, alternatives listed in the scale is treated as equal-distance and scored as successive integral. For example, score of 4, 3, 2, or 1 was given if the alternative "strongly agree", "agree", "disagree", or "strongly disagree" was chosen, respectively. This scoring approach, however, has been criticized on the grounds that it is too simplistic [2]. First, the assumption of equal-distance of adjacent alternative is questionable. Second, concerning the characteristic of variables, the descriptive terms applied in rating scales are linguistic variables rather than numerical variables. Consequently, utilizing fuzzy set theory (FST) to score psychological measurement seems feasible. However, in contrast with the many engineering studies discussing fuzzy set theory, only a few such works have been published in psychological measurement [3].

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 [5], 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) [7] by psychiatric doctors.

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.

Received 2007.05.20
Accepted 2007.07.29

2. METHODS

2.1 Subjects

The total sample used in this study consisted of participants recruited from two separate populations: (a) outpatients of a psychiatric clinic who were diagnosed as suffering from depression as the depressed sample, and (b) undergraduates as the non-depressed sample.

A total of 240 subjects were selected from outpatients who visit the psychiatric clinic at Taipei Municipal Hoping Hospital and were diagnosed as having depression symptoms. The self-reported instrument utilized in this study was administrated by the researcher while the severity of depression was diagnosed by a psychiatrist.

A total of 321 Taiwan undergraduates from Taiwan were recruited in this study as non-depressed sample.

2.2 Measures

Beck Depression Inventory–II (BDI–II) [5] was adopted as the measure in study. BDI, specifically developed to address all of the nine DSM–IV criteria for a major depressive episode, is a self-reported instrument for measuring the severity of depression in adolescents and adults through items showing varying degrees of the main cognitive, affective, and physiological aspects of clinical depression. Participants circle the number (0 to 3) associated with the item that best describes how they had felt over the past two weeks. According to Beck et al. [5], total BDI–II scores ranging from 0 to 13 represent normal to minimal depression, total scores from 14 to 19 are mild, total scores from 20 to 28 are moderate, and total scores from 29 to 63 are severe.

2.3 Scaling

This study proposed the Fuzzy Partial Credit Scaling (FPCS), combining fuzzy set theory and partial credit model, as an alternative scoring method. The following paragraph illustrated the procedures to complete FPCS.

The procedures for Fuzzy Partial Credit Scaling were as follows:

Step 1: Subjects are asked to choose and assign percentages on alternatives of items. The sum of assigned percentages in each item must be constrained to 100%.

Step 2: Calculate the traditional scoring according to the procedures mentioned above.

Step 3: Calculate "step parameters" (δ_{ij}) defined in PCM as shown in Figure 1. The PCM algorithm [6] is shown in Equation 1. The probability of person j scoring x on item i was expressed as:

$$P_{ijx}=\frac{\exp\sum_{k=0}^{x}(\theta_j-\delta_{ik})}{\sum_{h=0}^{m_i}\exp\sum_{k=0}^{h}(\theta_j-\delta_{ik})} \quad x=0,1,\ldots,m_i \tag{1}$$

Step 4: Fuzzify crisp data into fuzzy data by constructing triangle fuzzy numbers using step parameters estimated in Step 3.

We try to map linguistic variables, Alternatives 1 to 4, into corresponding reasonable normal fuzzy numbers $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, and $\tilde{D}$, with triangular membership functions $\mu_{\tilde{A}}$, $\mu_{\tilde{B}}$, $\mu_{\tilde{C}}$ and $\mu_{\tilde{D}}$. These membership functions are shown in Figure 1.

The x-axis represents ability, usually ranging from -3 to 3; while y-axis represents degree of membership, ranging from 0 to 1.

In Figure 2, we first find the "step parameters" (δ_{ij}) estimated by PCM. We propose that subject with ability located between -3 and "step parameter 1" (δ_{ij}) will choose Alternative 1. For this reason, the triangular fuzzy number $\tilde{A}$ = (-3, -3, δ_{i1}) with -3 and δ_{i1} being the lower and upper bounds, respectively, and -3 as the most likely value for $\tilde{A}$. In Figure 2, we draw a line segment from (-3, 1) to (δ_{i1}, 0) to characterize the membership of function of $\tilde{A}$.

Next, we propose that subject with ability located between "step parameter 1" (δ_{i1}) and "step parameter 2" (δ_{i2}) will choose Alternative 2 and the middle point between these two step parameters should receive the maximum degree of membership. Therefore, the triangular fuzzy number $\tilde{B}$ = (δ_{i1}, (δ_{i1} + δ_{i2})/2, δ_{i2}) with δ_{i1} and δ_{i2} being the lower and upper bounds, respectively, and (δ_{i1} + δ_{i2})/2 being the middle point which is the most likely value for $\tilde{B}$. In Figure 2, we draw a line segment from (0, δ_{i1}) to ((δ_{i1} + δ_{i2})/2, 1) to represent the left leg and another line segment from ((δ_{i1} + δ_{i2})/2, 1) to (δ_{i2} , 0) to represent the right leg of the triangular fuzzy number.

Likewise, we proposed $\tilde{C}$ = (δ_{i2} , (δ_{i2} + δ_{i3})/2, δ_{i3}) and $\tilde{D}$ = (δ_{i3} , 3, 3) to characterize the likelihood of Alternatives 3 and 4, respectively.

Step 5: Defuzzify fuzzy data into scalar using the center of gravity (COG) method. COG calculates the center of gravity of the support of the fuzzy number weighted by the membership grade.

The center of gravity of fuzzy set $\tilde{X}$ with membership function $\mu_{\tilde{A}}$, GR(X) =

$$\frac{\int_{-\infty}^{\infty} x\mu_x(x)dx}{\int_{-\infty}^{\infty}\mu_x(x)dx} \tag{2}$$

For a triangular fuzzy number $\tilde{X}$(a, b, c), GR(X) = (a+b+c)/3 [4].

Step 6: Calculate the fuzzy observed scores.

After defuzzification, we calculate the fuzzy observed

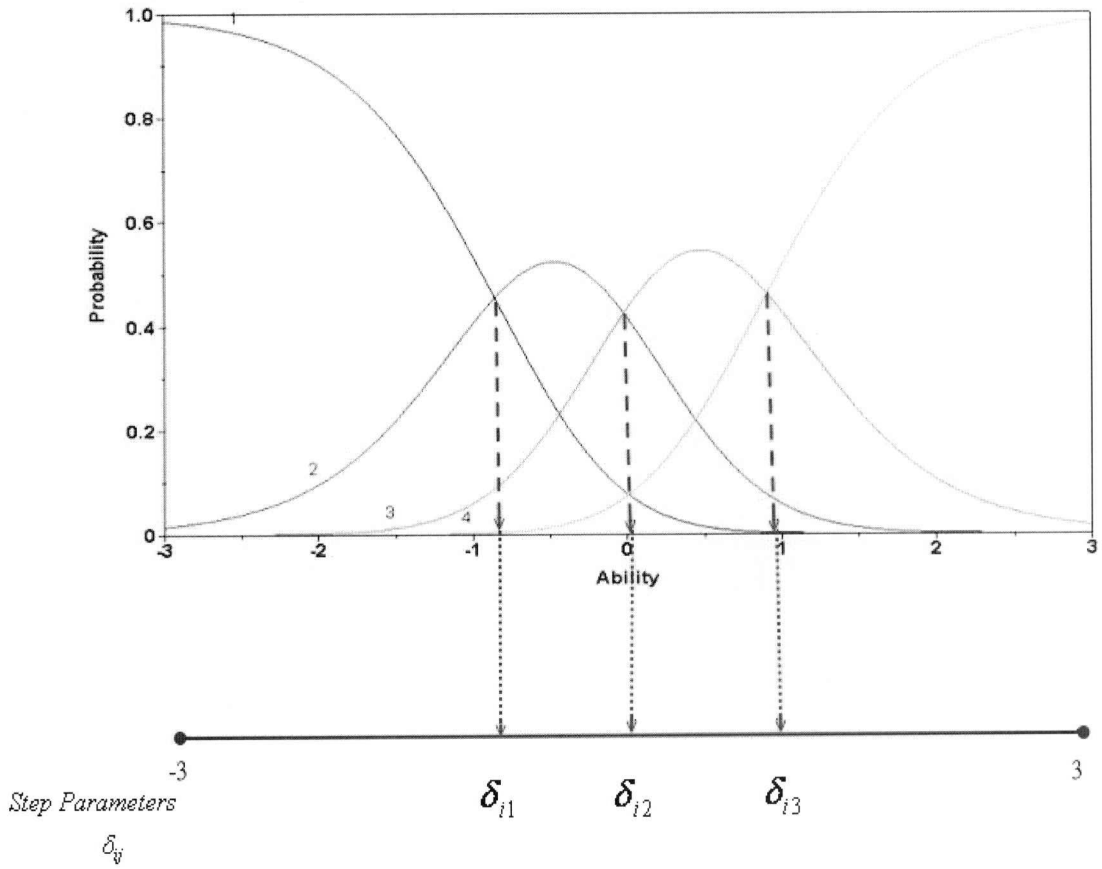


Figure 1: Calculations of Step Parameters

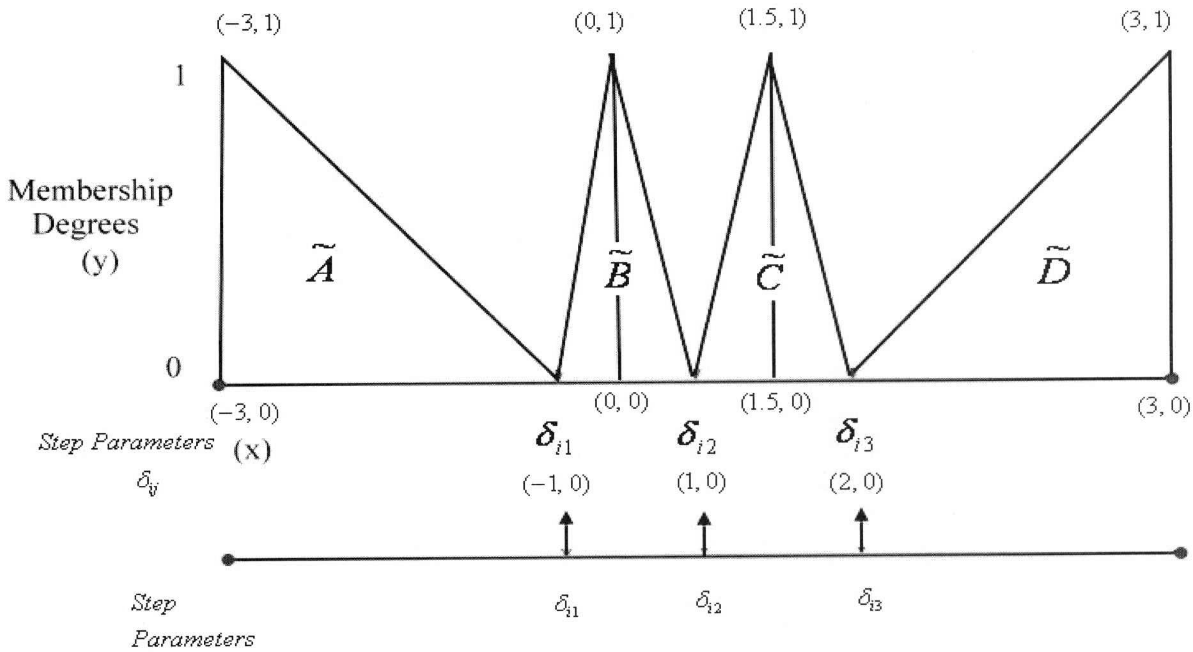


Figure 2: Constructions of Triangular Fuzzy Numbers

scores by multiplying the sum of membership degree of the set by its center of gravity.

A hypothesized example is shown in Table 1. In this table, an item of four alternatives with corresponding fuzzy sets $\tilde{A} \sim \tilde{C}$ is given. As noted previously, we estimate step parameters of PCM to construct fuzzy triangular numbers. Next, we calculate the center of gravity of these fuzzy numbers for defuzzfication. Finally, the fuzzy observed scores is obtained by weighing the sum of center of gravity by degree of membership.

2.4 Procedures

Depressed participants were recruited from Taipei Municipal Hoping Hospital. Those who were diagnosed as suffering from depression were asked to complete the BDI. The self-reported instrument was administered by the researcher one subject at a time. After completing the instrument, a short interview was given to the subject to collect their feedback about items in the Chinese version BDI and about the new "fuzzy" method to fill out a scale.

Non-depressed participants were recruited from undergraduates of pedagogy courses and they also completed the same self-reported instrument. To avoid contamination, undergraduates who meet cutoffs on BDI, 19 points according to traditional scoring, are excluded.

3. RESULTS AND DISCUSSION

The validity of an instrument is how well it measures what it purports to measure. Predictive validity was employed in this study to investigate the validity of FPCS. Predictive validity refers to the functional relationship between a predictor and a criterion event occurring before, during, and after the predictor is applied [2]. In this section, two different scoring schema, raw scores and FPCS, yielded two different predictors whereas suffering from depression as diagnosed by psychiatrist served as the criterion.

The total sample was separated into three groups: non-depression, depression with remission, and depression without remission. Discrimination analysis was used to predict these three group membership (non-depression, depression with remission, and depression without remission) from FPCS and raw scores, respectively. Classification accuracy was adopted as the criteria to evaluate the predictive validity of BDI scoring via FPCS and raw scores, respectively.

So far as FPCS was concerned, the results of discrimination analysis were listed in Table 2. The classification results presented in Table 2 showed that 80.7% of original grouped cases were correctly classified. Regarding the discrimination analysis results of raw score, both discrimination functions were statistically significant at $\alpha = .05$. However, only 71.2% of original grouped cases were correctly classified. That is, via FPCS, the probability of correct classification of severity of depression was raised from 71.2% to 80.7%.

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. 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.

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 [2]. Therefore, numbers or objects are often

Table 1: A Hypothetic Example of FPCS

	Membership Degrees	Step Parameters (δ_{ij})	center of gravity of fuzzy set (GR)	The fuzzy observed scores
$\tilde{A}$ (Alternative 1)	.8	Lower Bound= -3	((-3)+(-3)+(-1))/3= -2.333	**(.8 × -2.333) + (.2×2)= -1.4664**
$\tilde{B}$ (Alternative 2)	.2	δ_{i1} = -1	((-3)+(-2)+(-1))/3= -2	
$\tilde{C}$ (Alternative 3)	0	δ_{i2} = 0	((-1)+(-.5)+0)/3=-.5	
$\tilde{D}$ (Alternative 4)	0	δ_{i3} = 1	(1+3+3)/3=2.333	
		Upper Bound= 3		

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 [3, 8]

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 [9]. 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 [10-11]. 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. 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. It is also suggested that considerable cost concerning prevention and cure of depression might be reduced via FPCS.

REFERENCE

1. Yu, S. (2005). *Fuzzy Partial Credit Scaling: Applying Fuzzy Set Theory to Scoring Rating Scales*. Unpublished doctoral dissertation, National Chengchi University, Taipei, Taiwan.
2. Nunnally, J. C., & Berstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hall.
3. Yu, S., & Wu, B. (2007). Fuzzy Partial Credit Scaling: A New Approach to generate membership function for psychological measurement. *Quality and Quantity*, in press.
4. Zimmermann, H, J. (1996). *Fuzzy set theory and its applications*. Boston: Kluwer.
5. Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory–II*. San Antonio, TX: Psychological Corporation.
6. Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
7. American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (fourth edition). Washington DC: American Psychiatric Association.
8. Dubios, D., Ostasiewicz, W., & Prade, H.(2000). Fuzzy sets: History and basic notions. In D. Dubios & H. Prade (Eds.), *Fundamentals of fuzzy* sets (pp.21-124). Boston: Kluwer.
9. Ross, T. J., Sellers, K. M., & Booker, J. M. (2002). Considerations for Using fuzzy set theory and probability theory. In T. J. Ross., J. M. Booker, & W. J. Parkinson (Eds.), *Fuzzy logic and probability applications*. Philadelphia: SIAM.

Table 2 (a) Classification Results (FPCS)

		Predicted Group		
		1	2	3
Original Group	Non-depression (1)	96.6%	.0	3.4%
	Depression with Remission (2)	31.4%	.0	68.6%
	Depression without Remission (3)	28.8%	.0	71.2%

Note. 80.7% of original grouped cases correctly classified.

Table 2 (b) Classification Results (Raw Score)

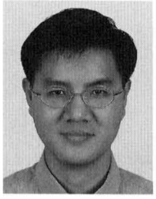
		Predicted Group		
		1	2	3
Original Group	Non-depression (1)	87.6%	.0	12.4%
	Depression with Remission (2)	37.1%	.0	62.9%
	Depression without Remission (3)	40.5%	.0	59.5%

Note. 71.2% of original grouped cases correctly classified.

10. Klir, G., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*. Englewood Cliffs, NJ: Prentice Hall.
11. Kosko, B. (1993). *Fuzzy thinking: The new science of fuzzy logic*. New York: Hyperion.

Sen-Chi Yu

Sen-chi Yu received his Ph. D degree in education from National Chengchi University – Taipei, Taiwan (2005). He is currently an assistant professor in the Teacher Education Center, HuaFan University, Taiwan. His current research interests focus on applied psychometric theory include structural equation modeling (SEM), item response theory (IRT), and fuzzy theory.



Yuan-horng Lin

Yuan-Horng Lin received his Ph. D degree in education from National Chengchi University – Taipei, Taiwan (1999). He is currently an associate professor in Department of Mathematics Education, National Taichung University, Taiwan. His current research interests include psychometric methodology, fuzzy theory, cognition diagnosis and knowledge structure analysis, and mathematics education.