Chapter Four

Results

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.

In Section I, data examinations were performed to ensure the legitimacy of the subsequent statistical analysis and feasibility of the proposed mathematical model. Some data screening procedures, including examinations of normality, outlier, missing data, were conducted. Moreover, since FPCS is derived from the Rasch model, some examinations about the assumptions of the Rasch model, including dimensionality and item fit analysis, were carried out.

In Section II, the procedures of constructing FPCS were demonstrated. The item steps, item-person map estimated via PCM and the defuzzfied triangle fuzzy numbers were given to calculate the fuzzy observed scores.

In Section III, the analytical results of Study One were demonstrated. The reliability of FPCS and that of raw scores were compared in this study.

In Section IV, the analytical results of Study Two were demonstrated. The validity of FPCS and that of raw scores were compared. Logistic regression and discrimination analysis were performed to provide predictive validity estimations.

In Section V, the results of Study Three were demonstrated. FCM, Wald's method, and k-means clustering were conducted to discuss whether fuzzy-based clustering method, FCM, is a more powerful tool than crisp-based ones to discover the structure of data.

I. Data Examinations

In this section, data examination procedures, including item analysis, dimensionality of BDI, examinations of normality, outliers, and missing data, were performed.

A. Item Analysis of BDI

IRT techniques provide a new approach to assessing the psychometric properties of attitude scales. The traditional CTT-based scale analysis or item analysis approaches had their limitations. For instance, item-test correlation coefficients are sample-dependent and violate local dependence (Embretson & Reise, 2000).

Concerning the limitations of traditional CTT-based item analysis and factor structure equality being violated, the IRT-based item analysis techniques were adopted to assess the psychometric properties of BDI in this study. Two chi-square ratios, the infit and outfit mean square statistics are indices most applied for scale analysis. The procedures for assessing infit and outfit were as follows.

Before estimates are applied to calibration and measurement, it is necessary to verify that the empirical data are suitable for the model. A useful approach to analyzing the fit is to compare the response of each person to each item with its expected value. When this comparison is summarized over the same person for an item, it indicates the overall validity of that item (Wright, 1999; Wright & Masters, 1982).

In PCM, the expected value of the response x_{ni} is given by (Wright & Masters, 1982)

$$E_{ni} = \sum_{k=0}^{m} k\,\pi_{nik} ,$$

where π_{nik} is person *n*'s modeled probability of responding in category *k* to item *i*. When the expected value E_{ni} is subtracted from the observed response x_{ni}, a score residual y_{ni} is obtained:

$$y_{ni} = x_{ni} - E_{ni}$$

Therefore, the variance of response x_{ni} is given by

$$W_{ni} = \sum_{k=0}^{m} (k - E_{ni})^2\,\pi_{nik} .$$

Subsequently, the residual y_{ni} can be standardized by

$$Z_{ni} = y_{ni} / (W_{ni})^{1/2} .$$

With the standardized residual obtained, one approach to summarizing the fit of an item is to square each of the standardized residual Z_{ni} and average these Z_{ni} over N persons. Consequently, for each item *i*, an unweighted mean square, the "outfit" mean square, is described by

$$U_i = \sum_{k=1}^{m} z_{ni}^{\ 2} / N \ .$$

Analogously, an unweighted mean square for each person *j*, is described by

$$U_j = \sum_{k=1}^{m} z_{ni}^{\ 2} / n \ .$$

Since U_i is sensitive to outliers, to reduce the impact of outliers, a weighted mean square can be computed by weighting the squared standardized residual $z_{ni}^{\ 2}$ by the information available about person *j* from item *i*. This is the basis of mean square (Masters & Wright, 1997).

For each item *i*, a weighted mean square (MNSQ), also called "infit" mean square, is computed as:

$$v_i = \sum_{n=1}^{N} z_{ni}^{\ 2} W_{ni} / \sum_{n=1}^{N} W_{ni} = \sum_{n=1}^{N} y_{ni}^{\ 2} / \sum_{n=1}^{N} W_{ni} \ .$$

For each person *j*, a weighted mean square (MNSQ) is calculated as

$$v_j = = \sum_{i=i}^{n} y_{ni}^{\ 2} / \sum_{n=1}^{N} W_{ni}$$

Infit and outfit statistics are reported as mean squares in their form of chi-square statistics divided by their degree of freedom, so that they have a ratio scale with expected value of +1 and are always positive. Some Rasch model computer programs also provide ZSTD statistics. ZSTD is the infit mean-square fit statistic t standardized to approximate a theoretical "unit normal", mean 0 and variance 1, distribution (Linacre, 2005).

An infit or outfit mean square value (MNSQ) of (1+ x) indicates (100x)% more variation between the observed and the expected value. For instance, an infit mean square value of 1.50 indicates 50% more variation in the empirical data than the Rasch model predicted. Likewise, an infit or outfit mean square value of (1- x) indicates (100x) % less variation between the observed and the expected value (Bond & Fox, 2001). In other words, MNSQ values substantially below 1 indicate dependency in data; while values substantially above 1 indicate noise. Values greater than 2 indicate that off-variable noise is greater than useful information whereas values less than 0.5 indicate that the data are overly predictable (Linacre, 2005).

Concerning the interpretation of the infit statistics, no absolute criteria have been proposed. It depends on the test situation and measurement context. Generally, reasonable MNSQ ranges for clinical observations are 0.5-1.7, while for rating scales, the ranges are 0.6-1.4. Moreover, the ZSTD is expected to less than 2 (approximately at $\alpha = .05$) or 2.58 (approximately at $\alpha = .01$). Regarding the application of ZSTD, it is only useful to salvage non-significant MNSQ>1.5, when the sample size is small or the test length is short (Bond & Fox, 2001; Linacre, 2005).

The mean-square or t standardized fit statistics are shown in Tables 4.1(a). As for the infit ZSTD, item 16 (changes in sleeping pattern) and item 21 (loss of interest in sex) exhibited worst fit with both t-values greater than 5.0. These most unfitted items were also reported by outpatients suffering from depression during the administration of BDI. Some teenagers and the elderly reported that they did not have sex lives. Moreover, almost all the outpatients suffered from insomnia and were taking sleeping pills. Therefore, even they were severely depressed, no sleeping problems were reported as long as they were on medication. These two items were deleted for their poor fit to the data.

After deleting items 16 and 21, the subsequent 19-item BDI was reanalyzed and the mean-square or t standardized fit statistics were shown in Tables 4.1(b). As seen in this table, the infit ZSTD of item 6 (punishment feelings), item 10 (crying), and item 18 (changes in appetite) exhibited worst fit with t-values all exceeding 2.58. Therefore, these three items were also deleted. The final version of BDI for subsequent analysis thus contained 16 items, as shown in Table 4.1(c).

Table 4.1(a) Item Fit of Original 21-item BDI

	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
Items				
1.Sadness	0.95	-0.68	0.83	-1.59
2.Pessimism	0.99	-0.10	1.03	0.24
3.Past Failure	1.11	1.53	1.39	2.37
4.Loss of Pleasure	0.88	-1.61	0.88	-1.06
5.Guilty Feelings	1.03	0.38	0.97	-0.37
6.Punishment Feelings	1.22	2.81	1.76	3.66
7.Self-Dislike	0.87	-1.91	0.67	-3.46
8.Self-Criticalness	0.98	-0.37	1.04	0.32
9.Suicide Thoughts	0.93	-0.81	0.81	-1.50
10.Crying	1.29	3.99	1.16	1.07
11.Agitation	0.9	-1.56	0.81	-1.64
12.Loss of Interest	0.8	-3.17	0.76	-2.50
13.Indecisiveness	0.87	-2.11	0.83	-1.85
14.Worthlessness	0.78	-3.24	0.69	-2.86
15.Loss of Energy	0.83	-2.88	0.83	-2.56
16.Changes in Sleeping Pattern	1.38	5.74*	1.52	6.71
17.Irritability	0.85	-2.45	0.72	-3.29
18.Changes in Appetite	1.17	2.39	1.49	5.01
19.Concentration Difficulty	0.99	-0.09	1.03	0.36
20.Sadness	1.01	0.09	1.00	0.00
21. Loss of Interest in Sex	1.4	5.02*	2.09	5.89

Note. * indicates Infit ZTSD > 5.0

Table 4.1(b) Item Fit of Modified 19-item BDI

	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
Items				
1.Sadness	0.97	-0.36	0.87	-1.25
2.Pessimism	1.02	0.31	1.06	0.58
3.Past Failure	1.17	2.37	1.44	2.71
4.Loss of Pleasure	0.92	-1.04	0.9	-0.87
5.Guilty Feelings	1.04	0.55	1.01	0.09
6.Punishment Feelings	1.25	3.08*	1.86	4.14
7.Self-Dislike	0.9	-1.55	0.7	-3.16
8.Self-Criticalness	1.01	0.21	1.05	0.4
9.Suicide Thoughts	0.97	-0.30	0.84	-1.28
10.Crying	1.29	4.00*	1.24	1.64
11.Agitation	0.89	-1.65	0.81	-1.7
12.Loss of Interest	0.84	-2.43	0.82	-1.96
13.Indecisiveness	0.91	-1.37	0.87	-1.44
14.Worthlessness	0.84	-2.25	0.77	-2.17
15.Loss of Energy	0.88	-2.05	0.88	-1.86
17.Irritability	0.89	-1.76	0.78	-2.67
18.Changes in Appetite	1.26	3.59*	1.8	7.78
19.Concentration Difficulty	1.04	0.56	1.13	1.35
20.Sadness	1.11	1.61	1.12	1.82

Note. * indicates Infit ZTSD > 2.58

Table 4.1(c) Item Fit of Modified 16-item BDI

	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
Items	1.06	0.76	1.08	0.74
1.Sadness	1.05	0.75	1.13	1.20
2.Pessimism	1.18	2.47	1.43	2.81
3.Past Failure	0.95	-0.69	0.93	-0.64
4.Loss of Pleasure	1.16	1.93	1.16	1.76
5.Guilty Feelings	0.91	-1.23	0.73	-2.95
7.Self-Dislike	1.11	1.51	1.20	1.55
8.Self-Criticalness	1.03	0.32	0.96	-0.29
9.Suicide Thoughts	0.97	-0.46	0.89	-0.98
11.Agitation	0.87	-1.90	0.87	-1.47
12.Loss of Interest	0.93	-1.13	0.93	-0.82
13.Indecisiveness	0.85	-2.05	0.77	-2.23
14.Worthlessness	0.91	-1.39	0.93	-1.09
15.Loss of Energy	0.95	-0.82	0.87	-1.61
17.Irritability	1.10	1.55	1.21	2.37
19.Concentration Difficulty	1.18	2.54	1.20	2.91
20.Sadness	1.06	0.76	1.08	0.74

B. Dimensionality of BDI

As with any model-based approach to interpreting psychological data, the advantages of IRT models are obtained in direct proportion to the degree to which the data are consistent with the assumptions. Most IRT models assume unidimensionality, that is, one major dominated latent trait variable is sufficient to explain the common variance among item responses (Embretson & Reise, 2000). Therefore, the dimensionality of BDI was evaluated.

Numerous heuristic approaches and statistical procedures for assessing dimensionality have been proposed, yet no definitive criteria have been established. Therefore, we adopted the "eigenvalue comparison criteria" (Hambleton & Swaminathan, 1985) and CFA (Embretson & Reise, 2000); both of which are most commonly used approaches.

In the first method, eigenvalues for test item intercorrelation matrices were plotted. Next, we looked for a dominant first factor and a high ratio of the first to second eigenvalue. The results of principle component analysis were presented in

Figure 4.2. The first eigenvalue was 10.44 while the second eigenvalue was only 1.09. Moreover, examining the scree plot revealed a significant “break” between the first and second eigenvalue. Given these findings, we concluded that there existed a dominant factor.

Concerning the second approach, second-order CFA was performed to confirm the factor structure of BDI. According to the original work by Beck et al. (1996), factor analysis of BDI performed on American psychiatric patients revealed two factors: somatic-affective and cognitive factors. The succeeding CFA was based on this factor structure.

Since this study consisted two heterogeneous samples: depressed and non-depressed populations, the equality of factor structure may be violated. So far as equality of factor structure was concerned, the researcher carried out second-order CFA on depressed and non-depressed samples respectively.

The analytical result of CFA based on depressed and non-depressed samples were illustrated in Figure 4.3. For depressed sample, the CFA model yielded good fit (Chi-Square= 120.80, p-value = 0.111 with df =103, CFI = 0.978, GFI = 0.941, RMSEA = 0.027). This analytical result of CFA indicated that, one higher order factor, depression, could be extracted from the two factors of BDI. That is, all items in BDI are measured by one single higher-order factor, depression. Therefore, the assumption of unidimensionality of BDI was supported by depressed sample.

However, for non-depressed sample, the CFA model yielded bad fit (Chi-Square= 876.14, p-value = 0.000 with df =103, CFI = 0.861, GFI = 0.745, RMSEA = 0.153). Therefore, the second-order factor model was not supported by non-depressed sample. The researcher argued that, BDI was designed for assessing the severity of clinical population; therefore, the factor structure of BDI derived from psychiatric patients exhibited bad data-model fit with non- psychiatric population.

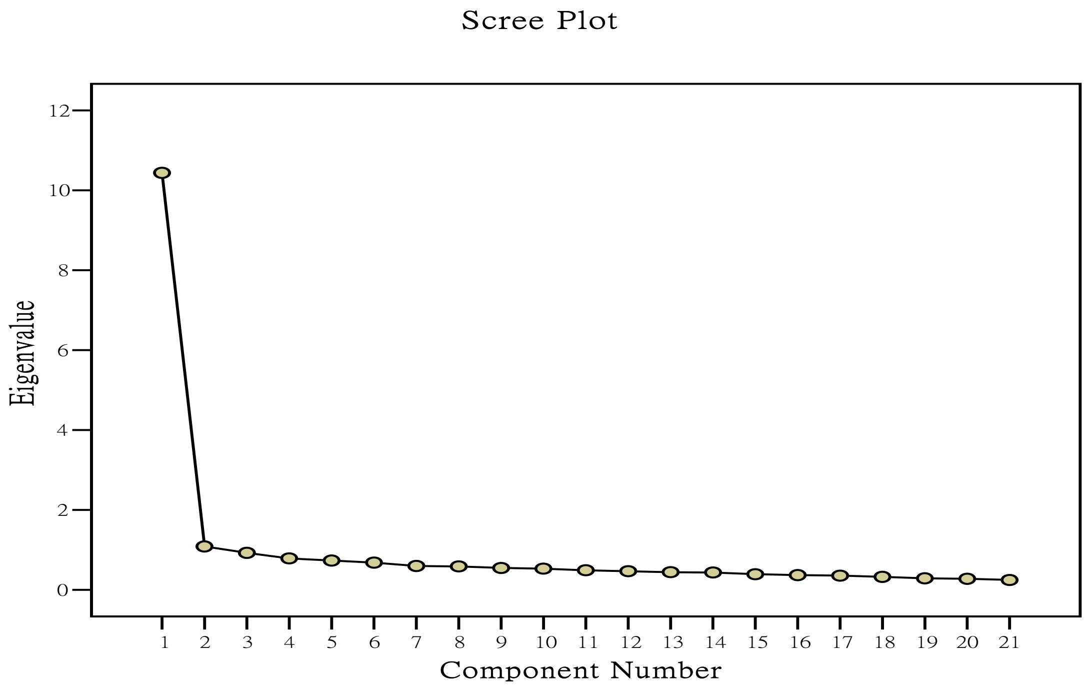

Figure 4.2 Scree Plot of Eigenvlues

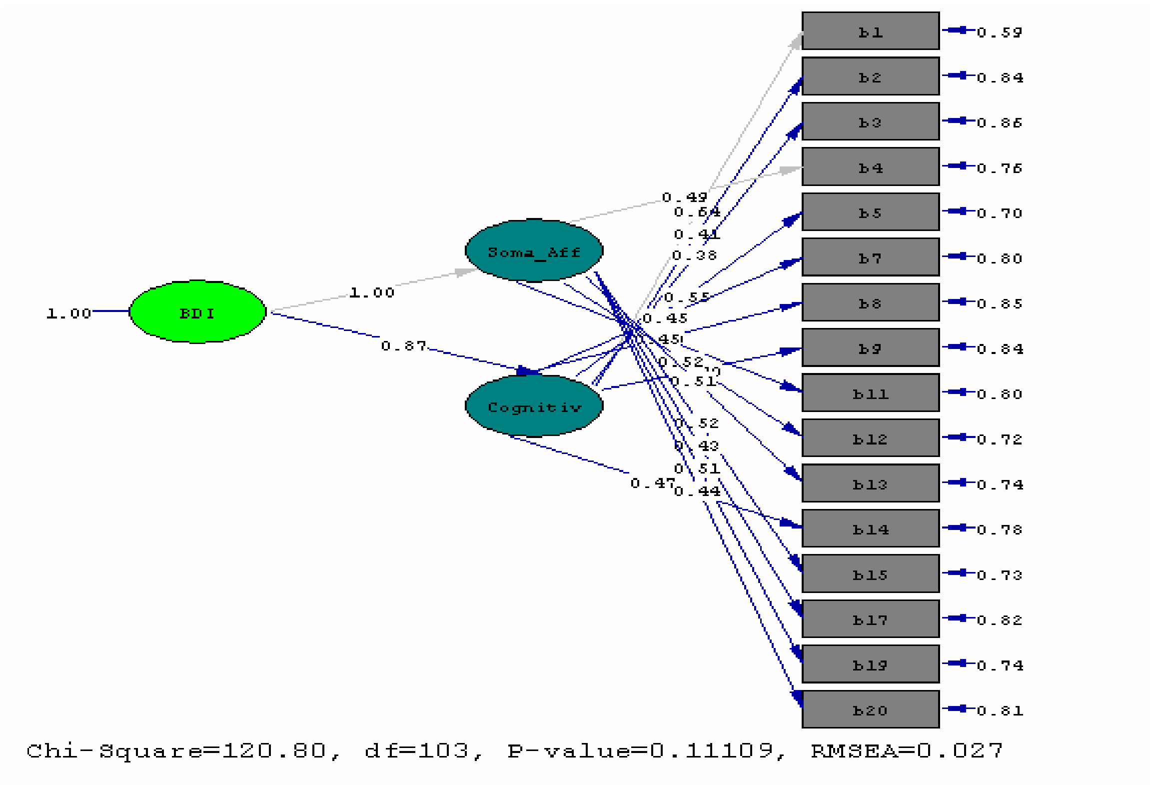

Figure 4.3 CFA of BDI Based on Depressed Sample

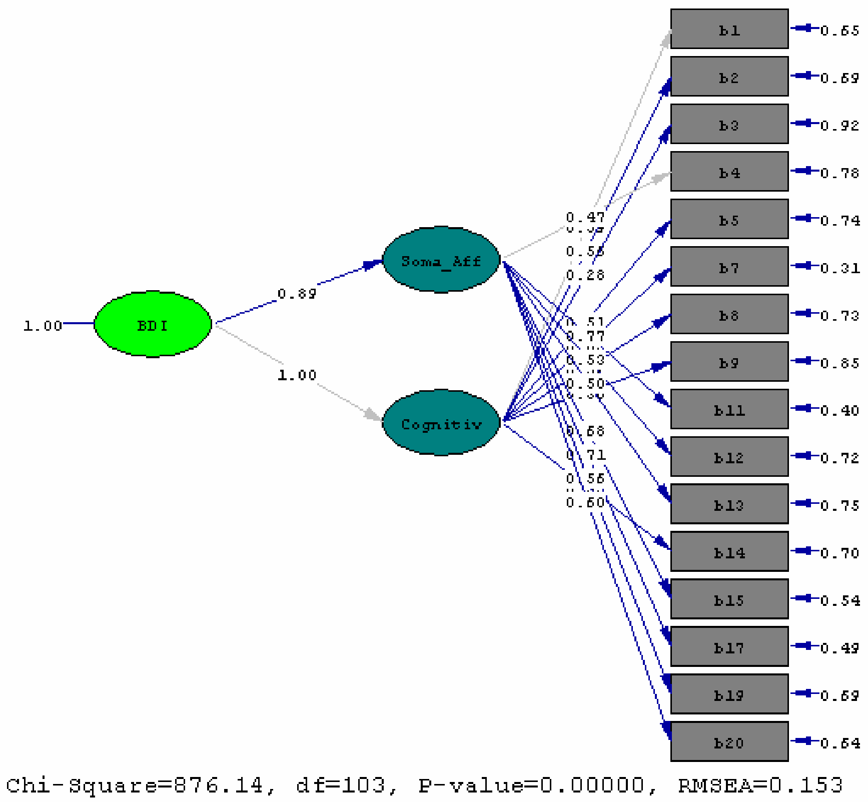

Figure 4.4 CFA of BDI Based on Non-depressed Sample

C. Data screening

Data screening can provide a basic understanding of the data and lead to better predictions. The normality, outlier, and missing data of the collected data were examined.

The analysis techniques utilized in this study included Rasch analysis, logistic regression, and SEM. In SEM, most estimation techniques and the chi-square test adopted in fitting function have multivariate normality assumed. Therefore, the normality assumption was examined in this section.

Since the succeeding analyses were performed based on the two subscales, somatic-affective and cognitive, of BDI. Statistical tests of normality of the two subscales based on FPCS and raw score were performed. The analytical results of testing the normality on the two different scaling approaches were given in Table 4.2. For depressed sample, the Kolmogorov-Smirnov test indicated that normality is verified in FPCS at $\alpha = .05$. However, concerning raw score, the assumption of normality was rejected.

So far as the non-depressed sample was concerned, normality of FPCS and raw score were violated. The departure from normal distribution may be attributed to the nature of the BDI. Since the BDI was designed to reflect the severity of depression symptoms among psychiatric patients, it discriminated well the levels of severity among patients, but performed poor for the general population.

As for the issue of missing data, only three subjects of the depressed sample failed to complete the questionnaire. The low drop-out rate might be due to the instrument being administered by the researcher one subject at a time. Moreover, a short interview was given to the patient when the scale was completed to collect comments about the BDI and the multiple-scoring schema. The most queried items were item 21 (loss of interest in sex) and item 16 (changes in sleeping pattern). Some teenage and elderly patients reported that they did not have sex lives, therefore they reported "no recent changes in interest in sex" or left this item blank. Moreover, almost all the outpatients had sleeping problems of various degrees and were taking sleeping pills as treatment. Therefore, even they were severely depressed, no sleeping problems were reported as long as they followed the doctor's prescription. The two most questioned items were also verified by the following Rasch item analysis.

Regarding the outliers of subjects' responses, since the instrument applied in this study was a closed form inventory in which each category to an item was assigned degree of membership (percentages) from 0% to 100% and the total of assigned percentages were restricted to 100%, any value out of this range was identified as outliers and eliminated from the data. So far, only one subject from the depressed sample, whose total endorsed percentages were not equal to 100%, was excluded.

Table 4.2 (a) Descriptive Statistics of Two Subscales of BDI: Depressive Sample

	FPCS		Raw Score	
Descriptive Statistics				
	Somatic-Affective Subscale	Cognitive Subscale	Somatic-Affective Subscale	Cognitive Subscale
Mean	-4.05	-3.99	7.94	7.12
Std. Deviation	7.65	6.51	5.57	5.54
Skewness	0.22	0.18	0.46	0.75
Kurtosis	-0.30	-0.45	-0.47	0.07
Normality Test				
Kolmogorov-Smirnov Statistic	0.05	0.05	0.11	0.10
p-value	.20	.20	.00	.00

Table 4.2 (b) Descriptive Statistics of Two Subscales of BDI: Non-depressive Sample

	FPCS		Raw Score	
Descriptive Statistics				
	Somatic-Affective Subscale	Cognitive Subscale	Somatic-Affective Subscale	Cognitive Subscale
Mean	-12.48	-11.52	2.85	1.72
Std. Deviation	4.66	3.68	3.33	2.61
Skewness	0.81	1.48	1.07	2.19
Kurtosis	0.01	3.26	0.18	6.69
Normality Test				
Kolmogorov-Smirnov Statistic	.104	.134	.198	.268
p-value	.00	.00	.00	.00

II. Construction of FPCS

Construction of FPCS involved several steps. First, step parameters defined in PCM were estimated. Second, triangular fuzzy numbers were constructed using step parameters to characterize distributions of each alternative value estimated from the PCM. Third, center of gravity (COG) method was applied to "de-fuzzify" the fuzzy number into a scalar to denote the fuzzy number value. The details of parameter estimation in FPCS were illustrated as follows.

Step parameters were estimated via IRT software Winsteps (Linacre, 2005) and listed in Table 4.3. As shown in this table, disordered categories existed in item 1, 3, 6, 8, 9, 10, 11, 14, and 21. Compared with the American sample (Beck et al., 1996), disordered categories existed in item 6, 9, 11, and 21. It indicated that these disordered categories are relatively rarely observed, i.e., occupies a narrow interval on the latent variable (Linacre, 2005). In other words, the relation between the symptoms mentioned in these items with depression may be U-shaped, not monotonic. Besides, the person-item map was listed in Table 4.4. As shown in this table, the most difficult step was 3rd step in "Guilty Feeling" (item 5) whose logit was about +2. Whereas the least difficulty step was 1st step in "Changes in Sleep Pattern" (item 16) whose logit was less than -2.

After step parameters were estimated, triangular fuzzy numbers were constructed using step parameters. The defuzzified of each alternative in BDI were listed in Table

4.5. The defuzzified scalar was used to calculate fuzzy observed score.

Table 4.3 Estimated Step Parameters in BDI

Items	1st Step	2nd Step	3rd Step
1.Sadness	-.69	1.02	-.33
2.Pessimism	-.82	.39	.42
3.Past Failure	-.39	-.64	1.03
4.Loss of Pleasure	-.86	.37	.49
5.Guilty Feelings	-1.63	.77	.85
6.Punishment Feelings	.48	.48	-.96
7.Self-Dislike	-.39	.17	.22
8.Self-Criticalness	-.28	-.44	.72
9.Suicide Thoughts	-.47	1.68	-1.21
10.Crying	-.07	.14	-.07
11.Agitation	-.15	-.34	.49
12.Loss of Interest	-.85	.18	.67
13.Indecisiveness	-.93	-.29	1.21
14.Worthlessness	-.40	.41	-.01
15.Loss of Energy	-1.42	.07	1.35
16.Changes in Sleeping Pattern	-1.50	.56	.94
17.Irritability	-.84	-.26	1.10
18.Changes in Appetite	-1.29	.36	.93
19.Concentration Difficulty	-.96	-.34	1.31
20.Sadness	-1.94	.56	1.38
21. Loss of Interest in Sex	-.25	-.31	.55

Table 4.4 Person-item Map

```
     PERSONS             MAP OF ITEMS - Thurstone thresholds
              <frequ>|<less>
2                    +
                  .  |
                  .  |                                            Guilty Feelings          .3
                     |                                            Indecisiveness           .3
                                                                  Irritability
                                                                  Tiredness
                    T|                                            Changes in Appetite      .3
                                                                  Loss of Energy
                                                                  Loss of Pleasure
                  .  |                                            Concentration Difficulty .3
                     |                                            Loss of Interest         .3
1                 .  +                                            Loss of Interest in Sex  .3
                                                                  Past Failure
                                                                  Self-Criticalness
                  .  |                   Guilty Feelings       .2 Pessimism                .3
                                                                  Self-Dislike
                                                                  Suicide Thoughts
                                                                  Worthlessness
                  .  |                   Loss of Pleasure      .2 Agitation                .3
                                          Suicide Thoughts           Sadness
                  #  |T
                .##  |                   Sadness               .2 Changes in Sleeping Pattern.3
                 .#  |S                  Changes in Appetite   .2
                                          Worthlessness
                .# S|                    Indecisiveness        .2 Punishment Feelings      .3
                                          Loss of Interest
                                          Loss of Interest in Sex
                                          Pessimism
                                          Self-Dislike
```

Table 4.4 Person-item Map (Continued)

```
0        .###  +M                               Irritability          .2  Crying                .3
                                                  Punishment Feelings
                                                  Tiredness
            .##  |                              Loss of Energy          .2
                                                  Past Failure
                                                  Self-Criticalness
             .#  |S Suicide Thoughts        .1  Agitation               .2
                                                  Concentration Difficulty
             ##  |                              Crying                  .2
            ###  |T Loss of Interest in Sex   .1  Changes in Sleeping Pattern.2
                      Punishment Feelings
        .######  |  Loss of Pleasure          .1
                      Past Failure
                      Sadness
                      Worthlessness
            .##  |  Pessimism                 .1
                      Self-Criticalness
  -1         .#  +  Agitation                 .1
                      Crying
                      Loss of Interest
                      Self-Dislike
             .#  |  Guilty Feelings           .1
                      Indecisiveness
           .## M|   Changes in Appetite       .1
                      Irritability
             .#  |  Concentration Difficulty  .1
             ##  |
            .##  |
             .#  |  Loss of Energy            .1
  -2        .##  +
                 |
            .##  |  Changes in Sleeping Pattern.1
                      Tiredness
            .##  |
                S|               <rare>|<more>
```

Table 4.5 Fuzzy Numbers Defuzzified by COG

	Alternative 1	Alternative 2	Alternative 3	Alternative 4
Items				
1.Sadness	-2.23	0.17	0.35	1.89
2.Pessimism	-2.27	-0.21	0.41	2.14
3.Past Failure	-2.13	-0.52	0.20	2.34
4.Loss of Pleasure	-2.29	-0.25	0.43	2.16
5.Guilty Feelings	-2.54	-0.43	0.81	2.28
6.Punishment Feelings	-1.84	0.48	-0.24	1.68
7.Self-Dislike	-2.13	-0.11	0.20	2.07
8.Self-Criticalness	-2.09	-0.36	0.14	2.24
9.Suicide Thoughts	-2.16	0.61	0.24	1.60
10.Crying	-2.02	0.04	0.04	1.98
11.Agitation	-2.05	-0.25	0.07	2.16
12.Loss of Interest	-2.28	-0.34	0.43	2.22
13.Indecisiveness	-2.31	-0.61	0.46	2.40
14.Worthlessness	-2.13	0.00	0.20	2.00
15.Loss of Energy	-2.47	-0.68	0.71	2.45
16.Changes in Sleeping Pattern	-2.50	-0.47	0.75	2.31
17.Irritability	-2.28	-0.55	0.42	2.37
18.Changes in Appetite	-2.43	-0.47	0.65	2.31
19.Concentration Difficulty	-2.32	-0.65	0.49	2.44
20.Sadness	-2.65	-0.69	0.97	2.46
21. Loss of Interest in Sex	-2.08	-0.28	0.12	2.18

III. Study One: The Reliability of FPCS

Reliability refers to how consistently the instrument measures whatever it was designed to measure. Some limitations exist in traditional CTT-based reliability estimations. First, the reliability of single-item cannot be estimated. Second, the concept of latent variable was not incorporated in the estimation of reliability. Third, observed variables cannot be indicators for more than one construct (Mueller, 1996).

Under the framework of CFA, the above-stated limitations had been improved. Bollen (1989) proposed that reliability can be estimated from the proportion of variance in an observed variable that is accounted for by latent constructs.

To compare the reliability of FPCS and raw scores, Cronbach's alpha coefficient and squared correlation between the observed and latent variables were adopted. The estimated reliability coefficients were listed in Table 4.6. As shown in this table, for each single item, FPCS exhibited higher reliability as well as the total 16-item BDI as a whole was concern.

Table 4.6 Reliability of BDI

	Raw Scores	FPCS
Reliability for 16-item BDI		
Cronbach's Alpha	.939	.949
Reliability for Single Item		
Items		
1.Sadness	.456	.555
2.Pessimism	.495	.572
3.Past Failure	.475	.546
4.Loss of Pleasure	.498	.549
5.Guilty Feelings	.362	.412
7.Self-Dislike	.587	.654
8.Self-Criticalness	.463	.497
9.Suicide Thoughts	.452	.494
11.Agitation	.546	.583
12.Loss of Interest	.608	.630
13.Indecisiveness	.559	.571
14.Worthlessness	.607	.672
15.Loss of Energy	.593	.675
17.Irritability	.510	.535
19.Concentration Difficulty	.523	.547
20.Sadness	.449	.543

IV. Study Two: The Validity of FPCS

The validity of an instrument is how well it measures what it purports to measure. Moreover, validity can also be viewed as the accuracy of specific inference made from its scores. Validity has been given three major meanings: (1) content validity: sampling from a universe of required content, (2) construct validity: measuring psychological attributes, and (3) predictive validity: establishing a statistical relationship with a particular criterion, that is, using an instrument to estimate some criterion behavior that is external to the instrument itself (Hopkins et al., 1990; Nunnally & Bernstein, 1994).

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 (Nunnally & Bernstein, 1994). 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.

A. The Prediction of the depressed and the non-depressed

As for the criterion, suffering from depression was a binary outcome and the predictors were continuous variables, therefore, logistic regression was applied to investigating the relation between scoring schemas and diagnosis of suffering from depression in this section. In the following section, discrimination analysis was employed to classify subjects according to their severity of depression symptoms. In this section, suffering from depression was coded 1 (depressed) or 0 (non-depressed), and the two factors of BDI, somatic-affective and cognitive factors, were adopted as predictors in logistic regression.

The goodness-of-fit of the fitted logistic regression models needs to be examined before it is accepted for use. Generally speaking, χ^2 and G^2 statistics are the most applied measures for testing the goodness-of-fit; large χ^2 or G^2 values provide evidence of lack of fit. However, χ^2 and G^2 for logistic regression models fitted without continuous or nearly-continuous predictors do not have approximate chi-squared distributions. These indices of fit are more properly applied when explanatory variables are categorical rather than continuous (Agresti, 1996). Since explanatory variables in this study were continuous, the Hosmer-Lemeshow statistic, which partitions predicted probabilities rather grouping explanatory variables, was employed to measure goodness-of-fit.

Concerning FPCS, the Hosmer-Lemeshow statistics equaled 11.684, with df = 8, p-value=0.166, indicating a good fit. Since the goodness-of-fit was appropriate, we proceed with the descriptions of parameter estimations.

The computer output of the estimated regression function was listed in Table 4.7(a). As shown, the estimated regression function was

$$\hat{y} = 2.098 + 0.091 \quad x_1 \quad + 0.188\, x_2 .$$

Where x_1 and x_2 denote somatic-affective factor and cognitive factor calculated from FPCS, respectively.

Regarding the odds of suffering from depression, for each unit increase in FPCS somatic-affective factor, the estimated odds multiplied by exp (0.091) = 1.095. Likewise, the estimated odds multiplied by exp (0.188) = 1.207 for each unit increase in FPCS cognitive factor. Moreover, the overall correctly predicted percentage was 77.2%, as shown in Table 4.7(b).

After reviewing the results of FPCS, the raw scores were described as follows. Concerning the goodness-of-fit, the Hosmer-Lemeshow statistics equaled 15.434, with df = 7, p-value = 0.031, indicating only decent fit.

The computer output of the estimated regression function was listed in Table 4.8(a). As shown in this Table, the estimated regression function was:

$$\hat{y} = -1.676 + 0.067 \quad x_1 \quad + 0.285 \quad x_2 .$$

Where x_1 and x_2 denote somatic-affective factor and cognitive factor calculated from raw scores, respectively.

Regarding the odds of suffering from depression, for each unit increase in FPCS somatic-affective factor, the estimated odds multiplied by exp (0.067) = 1.069. Likewise, the estimated odds multiplied by exp (0.285) = 1.330 for each unit increase in raw-score cognitive factor. Moreover, the overall correctly predicted percentage was 74.8%, as shown in Table 4.8(b).

These findings reveal that FPCS, compared with raw scores, yields better model fit and more accurate estimation for predicting depression. Since predictive validity refers to the functional relationship between a predictor and a criterion event, these findings provide evidence that, for BDI, FPCS yielded higher predictive validity than raw score. The analytical results showed that, via FPCS, the probability of correct classification of depressed and non-depressed was raised from 74.8% to 77.2%.

Table 4.7(a) Estimated Logistic Regression Coefficients (FPCS)

	B	S.E.	Wald Statistics	p-value
Somatic-Affective Factor	.091	.028	10.979	.001
Cognitive Factor	.188	.034	31.115	.000
Constant	2.098	.235	79.677	.000

Table 4.7 (b) Classification Table (FPCS)

		Predicted Group		
		Non-depressed	Depressed	Percentage Correct
Observed Group	Non-depressed	256	62	80.5
	Depressed	65	175	72.9
Overall Percentage				77.2

Table 4.8(a) Estimated Logistic Regression Coefficients (Raw Score)

	B	S.E.	Wald Statistics	p-value
Somatic-Affective Factor	.067	.035	3.640	.056
Cognitive Factor	.285	.044	41.375	.000
Constant	-1.676	.160	109.526	.000

Table 4.8 (b) Classification Table (Raw Score)

		Predicted Group		
		Non-depressed	Depressed	Percentage Correct
Observed Group	Non-depressed	263	56	82.4
	Depressed	85	155	64.6
Overall Percentage				74.8

B. The Classification of Severity of Depression Symptoms

Validity of a measure can be view as the accuracy of specific inference made from its scores. To investigate the validity of FPCS, predictive validity was applied to classify total sample into the depressed and the nondepressed according to their responses to BDI in previous section. While in this section, the researcher discussed the validity of FPCS by classifying the severity of depression symptoms. According to

DSM-IV, the severity of major depressive episode, also called Fifth Digit Severity Code for Major Depressive Episode, was classified into seven categories: mild, moderate, severe with psychotic features, severe without psychotic features, partial remission, full remission, and unspecified (APA, 1994).

Concerning the above-mentioned classification, mild episode refers to symptoms barely meet criteria for major depression and result in little distress or interference with the patient's ability to work, study or socialize. Episode of severe without psychotic features means the number of symptoms well exceeds the minimum for diagnosis, and they markedly interfere with patient's work, social or personal functioning. Episode of severe with psychotic features indicates that the patient has delusions or hallucinations, which may be mood-congruent or mood-incongruent. In Partial Remission refers to patients who formerly met full criteria for Major Depressive Episode and now either have fewer than five symptoms or have had no symptoms for less than two months. In full remission refers to the patient has had no material evidence of Major Depressive Episode during the past 2 months (APA, 1994; PsychNet-UK, 2005).

To conclude from the foregoing paragraph, the classification of severity or episode of depression involved to what extent the patient meet the criteria, the time course, the initial diagnose, etc. That is, the severity of depression symptom is not the only classification criteria for the seven categories therefore these are not linearity separable. Therefore, it was not easy to classify the patient into the seven categories according to their response on depression scale. A simple but better classification was to partition patients into two groups only: remission and non-remission. Therefore, the total sample was separated into three groups: non-depression, depression with remission, and depression without remission.

In this section, 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 4.9 and 4.10. The classification results presented in Table 4.10 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%.

Table 4.9 Group Statistics

		FPCS		Raw score	
Classified Group		Mean	S.D.	Mean	S.D.
Non-depressive	Somatic-Affective Factor	-13.09	4.03	2.41	2.88
	Cognitive Factor	-12.10	2.82	1.28	1.87
Depression with remission	Somatic-Affective Factor	-7.00	8.40	6.37	5.83
	Cognitive Factor	-4.80	6.54	6.89	5.60
Depression without remission	Somatic-Affective Factor	-3.55	7.42	8.20	5.50
	Cognitive Factor	-3.85	6.51	7.16	5.54
Total	Somatic-Affective Factor	-9.06	7.43	4.87	5.10
	Cognitive Factor	-8.48	6.29	3.88	4.90

Table 4.10(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 4.10 (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.

V. Study Three: Fuzzy C-means Clustering

Pattern recognition, data analysis and cluster analysis, are often applied synonymously to identify structures in data (Bezdek, 1981; Hoppner et al., 1999; Zimmermann, 1996). In data analysis, objects are described by attributes to find structures or information in these data. The methodologies applied for recognition schemas include linear classification, statistical (probability) approaches, fuzzy set theory (possibility approaches), perceptions (neural networks) and knowledge-based classification based on artificial intelligence (Lin & Lee, 1999).

Despite the different terminology adopted in different disciplines, all clustering methods classify data according to "natural relationships" (Hair, Anderson, Tatham, & Black, 1998). Conventional crisp clustering methods assign each object into one exact group; that is, the group membership is 0 or 1. These methods are inappropriate, however, when some subjects are exactly at the intersection of two groups, or when groups overlap each other. That is, crisp clustering methods fail to deal with when some subjects belong equally to two groups. Conversely, fuzzy cluster analysis dispenses with unambiguous mapping of data by class and cluster, and instead calculates degrees of membership to specify the extent to which data belong to clusters (Hoppner et al., 1999).

Clustering analysis is an "unsupervised" technique, that is, no prior information was given to judge about what the output should be or whether it is correct. That is, group membership is unknown in cluster analysis, but is known in discrimination analysis. However, to compare the difference between crisp-based and fuzzy-based cluster analysis, the original group membership (non-depression, depression with remission, and depression without remission) were used as the criteria to evaluate which the cluster technique could accurately discover the structure of data. To attain this goal, FCM, Wald's method, and the *k*-means clustering method were utilized. The associations between classification results and original group membership resulting from different methods were compared.

First, cluster validity indices were applied to determine the optimal cluster number (c) and exponential weight (m) in FCM. Cluster validity represents the degree to which the final partition of a cluster algorithm approximates the real or hypothesized structure of a data set. A structure with many points concentrated around the cluster centers indicates a cluster result with high cluster validity. That is, the membership matrix of the final partition generated by the FCM algorithm is crisper (unfuzzier) than others(Zimmermann, 1996).

In this study, two indices, partition coefficient (PC) and partition entropy (PE) were employed as the measure of cluster validity. PE is expressed as PC =

$$\sum_{k=1}^{n}\sum_{i=1}^{c}\frac{(\mu_{ik})^2}{n} \quad .$$

Where μ_{ik} denotes the degree to which subject k belongs to cluster i, and c denotes the cluster number. The value of PC ranges from 1/c to 1. In the case of crisp partition, the maximum value PC = 1 is obtained. If the partitions contain no information, i.e., each datum is assigned to each cluster at the same degree, then the minimum value of PC result. The maximum PC is found for the partition with the "most unambiguous" assignment (Hoppner et al., 1999; Zimmermann, 1996).

Another cluster validity index, partition entropy, is related to Shannon's information. Shannon defined I_k, the information contents of a single event k with probability μ_k, as $I_k = -\ln(\mu_k)$, and the mean information content of a source as the entropy H = $-\sum_{k \in K} \mu_k \ln(\mu_k)$. Partition entropy is expressed as

$$\text{PE} = -\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c}\mu_{ik}\ln(\mu_{ik}).$$

Where μ_{ik} denotes the degree to which subject k belongs to cluster I, and c the cluster number. The ranges for PE are $0 \leqq \text{PE} \leqq \ln(c)$. Regarding crisp clustering, since the maximum information was obtained, the entropy is zero. By contrast, the entropy rises to a maximum when its membership distribution is uniform. A partition with low entropy is preferred to one with high entropy (Hoppner et al., 1999; Zimmermann, 1996).

The validity indices were computed by f-cut fuzzy partition software (Lin, 2003), and are shown in Table 4.11. The optimal cluster number was searched using maximum PE and minimum PC. As presented in Table 4.11, the optimal cluster number is 3 with exponential weight 1.25. Therefore, FCM apply with cluster number = 3 for the proceeding analysis.

To identify the cluster technique which could discover the data structure most accurately, cluster number = 3 was assigned to FCM, Wald's method, and k-means method. To compute the association between original and classified membership, classified crisp membership in FCM was modified by assigning 0-or-1 membership to the cluster with the highest membership. Kendall's τ coefficient, which measures the relationships among variables and rank orders, was applied to measure the association between original and classified membership. The results of association analysis are shown in Table 4.12.

As presented in Table 4.12, the Kendall's τ coefficient of FCM was .549, that of Wald's method was .316, and that of k-means was .395 ($p<.001$). The analytical results demonstrate that FCM exhibited a higher association between the original and classified membership than did Wald's and k-means methods. That is, FCM identified the data structure more accurately than the two crisp clustering methods.

Table 4.11 Summary of Cluster Validity

			Cluster Number (c)			
		Index	C=2	C=3	C=4	C=5
Exponential Weight(m)	m = 1.25	PE	.955	.965*	.940	.932
		PC	.080	.061**	.103	.117

Note. * indicated optimal cluster number according to PE
** indicated optimal cluster number according to PC

Table 4.12(a) Classification Results (FCM)

	Classified Group (FCM)			Total
Original Group	1	2	3	
Nondepressed (1)	229	82	8	319
Depression with Remission (2)	8	17	10	35
Depression without Remission (3)	35	86	84	205
Total	272	185	102	559
Kendall's τ coefficient	.549 (p<.001)			

Table 4.12(b) Classification Results (Wald's Method)

	Classified Group (Wald)			Total
Original Group	1	2	3	
Nondepressed (1)	265	5	49	319
Depression with Remission (2)	23	7	5	35
Depression without Remission (3)	93	60	52	205
Total	381	72	106	559
Kendall's τ coefficient	.316 (p<.001)			

Table 4.12(c) Classification Results (K-means)

	Classified Group (K-means)			Total
	1	2	3	
Nondepressed (1)	244	4	71	319
Depression with Remission (2)	10	8	17	35
Depression without Remission (3)	52	60	93	205
Total	306	72	181	559
Kendall's τ coefficient	.395 (p<.001)			