CHAPTER V

RESULTS

Descriptive ANOVA was used for analysis because of the following reasons: a) In Case I, where sample size was N=1000, it was evident that significance tests of mean differences would be redundant due to very low variation across replicated samples (see Table V-1 and Table V-2); b) In Case II and Case III, where sample sizes were N=100 and N=30 respectively, a significance test with ANOVA seemed inappropriate because the homogeneity assumption was violated (see Table V-3 to Table V-6); and c) Practical significance was of greater concern than was statistical significance in the current situation. The following analyses would focus on systematic patterns and practical significance. Practical significance was roughly judged by the ratio of each mean difference to the standard deviation of the combined sample. For being judged as "practical significance" a mean difference of at least one standard deviation was considered. When the two standard deviations involved in comparing means were highly heterogeneous, a more conservative conclusion was drawn.

The patterns of results were highly similar across the two assessing criteria, Pearson correlations and RMSDs. Therefore, most conclusions were based on both criteria.

Case I: N=1000

For each experimental condition in this case, the mean of five replicated correlations as well as of five replicated RMSDs between true and recovered person parameters were presented in Table V-1 and Table V-2 respectively. 1) Generally speaking, IRT-GRM performed best among the four procedures in terms of recovering the latent trait parameters. However, a few interactions between the recovering procedures and the distributions of item responses could also be observed. First, compared to FA-PL and FA-PR, IRT-GRM performed just as well as did the two FA procedures when the item responses were normally distributed but it tended to outperform them when the distributions of item responses became skewed. The advantage of IRT-GRM was especially evident when the distributions

Table V-1

Correlations between True and Recovered Person Parameters for Case I (N=1000)

#items	Dists.	IRT-GRM	FA-PR	FA-PL	SSI
12	Normally Distributed	.95 (.00)	.95 (.00)	.95 (.00)	.93 (.00)
	Moderately Skewed	.94 (.00)	.92 (.00)	.92 (.00)	.91 (.00)
	Highly Skewed	.88 (.00)	.81 (.01)	.80 (.01)	.81 (.01)
	Differentially Skewed	.93 (.00)	.93 (.00)	.93 (.01)	.91 (.00)
24	Normally Distributed	.98 (.00)	.98 (.00)	.97 (.00)	.96 (.00)
	Moderately Skewed	.97 (.00)	.94 (.00)	.94 (.00)	.94 (.00)
	Highly Skewed	.92 (.00)	.83 (.00)	.83 (.01)	.85 (.01)
	Differentially Skewed	.97 (.00)	.96 (.00)	.96 (.00)	.95 (.00)

Note: Each correlation in the entry is a mean of five correlations from five replications. The number in the parentheses is the standard deviation of the five replicated correlations.

Table V-2

RMSDs between True and Recovered Person Parameters for Case I (N=1000)

#items	Dists.	IRT-GRM	FA-PR	FA-PL	SSI
12	Normally Distributed	.31 (.01)	.31 (.01)	.31 (.01)	.36 (.01)
	Moderately Skewed	.33 (.01)	.39 (.00)	.40 (.01)	.43 (.01)
	Highly Skewed	.49 (.01)	.62 (.01)	.62 (.01)	.61 (.01)
	Differentially Skewed	.36 (.01)	.38 (.01)	.39 (.01)	.43 (.01)
24	Normally Distributed	.22 (.01)	.22 (.01)	.22 (.01)	.26 (.01)
	Moderately Skewed	.24 (.01)	.34 (.01)	.34 (.01)	.35 (.00)
	Highly Skewed	.39 (.01)	.58 (.01)	.59 (.01)	.55 (.01)
	Differentially Skewed	.26 (.01)	.27 (.01)	.28 (.01)	.31 (.01)

Note: Each quantity in the entry is a mean of five RMSDs from five replications. The number in the parentheses is the standard deviation of the five replicated RMSDs.

of item responses were highly skewed. Second, compared to the common SSI procedure, IRT-GRM performed slightly better when the distributions of item responses were normal, moderately skewed, or differentially skewed but much better when the distributions were highly skewed. It seemed that IRT-GRM was more robust against skewness than were the other three procedures involved. These results were true across the two conditions of test length.

2) FA-PR performed as well as did FA-PL. This finding was true across all experimental conditions of test length and of response distributions.

3) Both FA-PL and FA-PR were only slightly better than SSI when item responses were normally distributed. This advantage of factor analysis became even smaller when distributions of item responses were moderately or differentially skewed. Moreover, both FA-PL and FA-PR performed slightly worse than SSI when the distributions of item responses were highly skewed. It seemed that factor analysis was more sensitive to the disturbance of skewness than was SSI. These results were true across the conditions of test length.

4) It is not surprising that both test length and distributions of item responses had a main effect. All four procedures performed better in the condition of 24 items than in the condition of 12 items. They performed best when the distributions of item responses were normal, next best when the distributions were moderately or differentially skewed, and worst when the distributions were highly skewed.

5) All recovering procedures were fairly stable across replicated samples based on the fact that all standard deviations of the five correlations/RMSDs were less than or equal to 0.01.

<u>Case II: N=100</u>

For each experimental condition in this case, the mean of five replicated correlations as well as of five replicated RMSDs between true and recovered person parameters were presented in Table V-3 and Table V-4, respectively.

1) Generally speaking, IRT-GRM performed best among the seven procedures in terms of recovering the latent trait parameters. This conclusion depended

somewhat on the condition of item response distributions. When the item responses were normally distributed, IRT-GRM performed just as well as did FA-PR, FA-PL, WMDU, and SSI but much better than did EMDU and IMDU. However, it outperformed all other six procedures when the distributions of item responses were moderately skewed or highly skewed. When the distributions of item responses were differentially skewed, IRT-GRM performed slightly better than did FA-PR, FA-PL and SSI. It seemed that IRT-GRM was more robust against skewness than were the other six procedures. These results were even clearer when the number of items was larger and when RMSDs instead of correlations were assessed.

2) Generally speaking, FA-PR, FA-PL and SSI had similar performances across most conditions. However, a slight tendency was observed: although FA-PR and FA-PL performed as well as or even slightly better than did SSI when the distributions of item responses were normal or moderately skewed, they performed worse than did SSI when the distributions were highly skewed. These results implied that both

Table V-3

Correlations between True and Recovered Person Parameters for Case II (N=100)

Dist.	IRT-GRM	FA-PR	FA-PL	SSI	WMDU	IMDU	EMDU
12 items							
Norm. Dist.	.95 (.01)	.95 (.01)	.95 (.01)	.95 (.01)	.94 (.01)	.84 (.06)	.56 (.31)
Moder. Skewed	.95 (.01)	.92 (.01)	.92 (.01)	.91 (.02)	.92 (.01)	.86 (.03)	.76 (.11)
Highly Skewed	.89 (.02)	.81 (.04)	.79 (.06)	.82 (.03)	.81 (.04)	.76 (.06)	.76 (.08)
Diff. Skewed	.92 (.02)	.91 (.02)	.92 (.02)	.90 (.03)	.83 (.02)	.83 (.02)	.56 (.10)
24 items							
Norm. Dist.	.97 (.00)	.97 (.00)	.97 (.00)	.96 (.00)	.97 (.00)	.93 (.02)	.71 (.29)
Moder. Skewed	.96 (.00)	.94 (.01)	.94 (.01)	.94 (.01)	.94 (.01)	.83 (.10)	.81 (.02)
Highly Skewed	.90 (.01)	.82 (.03)	.84 (.03)	.85 (.03)	.81 (.05)	.83 (.07)	.64 (.18)
Diff. Skewed	.97 (.01)	.96 (.01)	.95 (.03)	.95 (.01)	.89 (.03)	.92 (.01)	.80 (.02)

Note: Each correlation in the entry is a mean of five correlations from five replications. The number in the parentheses is the standard deviation of the five replicated correlations.

Table V-4

RMSDs between True and Recovered Person Parameters for Case II (N=100)

Dist.	IRT-GRM	FA-PR	FA-PL	SSI	WMDU	IMDU	EMDU
12 items							
Norm. Dist.	.31 (.03)	.32 (.04)	.30 (.02)	.33 (.02)	.33 (.02)	.55 (.11)	.90 (.30)
Moder. Skewed	.33 (.02)	.39 (.03)	.39 (.03)	.42 (.04)	.39 (.03)	.52 (.06)	.67 (.15)
Highly Skewed	.46 (.04)	.61 (.07)	.64 (.09)	.59 (.06)	.60 (.06)	.69 (.09)	.68 (.11)
Diff. Skewed	.39 (.06)	.42 (.06)	.40 (.06)	.43 (.06)	.58 (.04)	.58 (.04)	.93 (.11)
24 items							
Norm. Dist.	.24 (.01)	.23 (.01)	.25 (.01)	.27 (.01)	.25 (.02)	.38 (.04)	.71 (.32)
Moder. Skewed	.26 (.01)	.35 (.02)	.36 (.02)	.35 (.02)	.35 (.03)	.57 (.15)	.62 (.04)
Highly Skewed	.44 (.03)	.59 (.05)	.56 (.05)	.55 (.05)	.62 (.07)	.56 (.13)	.83 (.20)
Diff. Skewed	.26 (.03)	.28 (.03)	.31 (.08)	.30 (.03)	.46 (.07)	.40 (.03)	.63 (.04)

Note: Each quantity in the entry is a mean of five RMSDs from five replications. The number in the parentheses is the standard deviation of the five replicated RMSDs.

FA-PL and FA-PR might be more severely disturbed by a high degree of skewness than was SSI.

3) WMDU performed as well as did IRT-GRM, FA-PR, FA-PL, and SSI when distributions of item responses were normal. It performed worse than IRT-GRM when distributions of item responses were moderately or highly skewed. It performed much worse than IRT-GRM, FA-PR, FA-PL, and SSI when item responses were differentially skewed.

4) Generally, the IMDU procedure performed worse than the traditional SSI procedure but better than EMDU in almost all conditions. It performed worse than did WMDU in most conditions except when item response distributions were differentially or highly skewed. If results from Case III (in Table V-5 and Table V-6) were also taken into consideration, it became clear that WMDU was generally better than IMDU.

5) EMDU performed worst among the seven procedures in almost all conditions.

6) With the exception of the two classical unfolding procedures (IMDU and EMDU), all procedures performed best when item responses were normally

distributed and worst when distributions were highly skewed.

7) Regardless of conditions of item response distributions, increasing test length improved estimation for all procedures except the two classical unfolding models.

8) In most conditions, the two classical unfolding models were less stable than the other five procedures because they occasionally produced inappropriate solutions which completely failed to recover the true person parameters.

Case III: N=30.

For each experimental condition in this case, the mean of five replicated correlations as well as of five replicated RMSDs between true and recovered person parameters were presented in Table V-5 and Table V-6 respectively.

1) Generally speaking, the common SSI procedure performed as well as or better than all three unfolding procedures in all conditions of test length and response distributions.

2) WMDU performed better than the two classical unfolding procedures, EMDU and IMDU, in all conditions.

3) IMDU performed better than did EMDU when the distributions of item responses were normal or differentially skewed. EMDU outperformed IMDU when distributions of item responses were highly skewed. The relative merits of IMDU and EMDU were not clear when distributions of item responses were moderately skewed.

4) Both of the SSI and the WMDU procedures had more stable performances across replications than did the two classical unfolding procedures. The latter two procedures were unstable because they sometimes produced inappropriate solutions which completely failed to recover the true person parameters of a few samples.

5) Both SSI and WMDU performed best when item responses were normally distributed and worst when item responses were highly skewed. However, SSI performed better when item responses were differentially skewed than when item responses were moderately skewed, while WMDU had the opposite

Table V-5

Correlations between True and Recovered Person Parameters for Case III (N=30)

#items	Dists.	SSI	WMDU	IMDU	EMDU
12	Normally Distributed	.93 (.02)	.92 (.05)	.63 (.24)	.56 (.22)
	Moderately Skewed	.89 (.05)	.88 (.05)	.68 (.25)	.64 (.11)
	Highly Skewed	.79 (.08)	.76 (.04)	.50 (.15)	.71 (.07)
	Differentially Skewed	.91 (.03)	.84 (.07)	.67 (.37)	.55 (.07)
24	Normally Distributed	.96 (.01)	.96 (.02)	.89 (.06)	.74 (.09)
	Moderately Skewed	.94 (.01)	.92 (.02)	.56 (.19)	.75 (.09)
	Highly Skewed	.87 (.04)	.81 (.04)	.44 (.26)	.69 (.05)
	Differentially Skewed	.94 (.02)	.91 (.02)	.88 (.09)	.79 (.05)

Note: Each quantity in the entry is a mean of five RMSD from five replications. The number in the parentheses is the standard deviation of the five replicated RMSD.

Table V-6

RMSDs between True and Recovered Person Parameters

for Case III (N=30)

#items	Dists.	SSI	WMDU	IMDU	EMDU
12	Normally Distributed	.13 (.03)	.39 (.10)	.80 (.26)	.90 (.22)
	Moderately Skewed	.21 (.10)	.47 (.09)	.75 (.28)	.82 (.12)
	Highly Skewed	.40 (.16)	.68 (.06)	.97 (.14)	.74 (.09)
	Differentially Skewed	.18 (.06)	.54 (.12)	.72 (.38)	.93 (.07)
24	Normally Distributed	.07 (.03)	.28 (.08)	.46 (.13)	.70 (.12)
	Moderately Skewed	.12 (.03)	.38 (.06)	.91 (.19)	.69 (.14)
	Highly Skewed	.26 (.07)	.60 (.06)	1.03 (.23)	.77 (.06)
	Differentially Skewed	.11 (.04)	.41 (.06)	.46 (.16)	.63 (.07)

Note: Each quantity in the entry is a mean of five RMSDs from five replications. The number in the parentheses is the standard deviation of the five replicated RMSDs.

pattern of performances.

6) The above outcomes became clearer when number of items became larger.

Conclusions across Cases

1) From the best to the worst, the general performances of the seven procedures could be ordered as follows: a) IRT-GRM, b) FA-PL and FA-PR, c) SSI, d) WMDU, and e) IMDU and EMDU.

2) IRT-GRM was more robust against skewness than were FA-PL, FA-PR, and SSI even when sample size was as small as N=100.

3) FA-PL and FA-PR were competitive with IRT-GRM only when item responses were normally distributed.

4) FA-PL and FA-PR performed equally well across almost all conditions.

5) The common SSI practice might be slightly worse than the two FA procedures when item responses were normally distributed, but it was better than them when item responses were highly skewed.

6) The WMDU procedure was generally not a better alternative to the common SSI practice. It

performed as well as did SSI only when item responses were normally distributed or moderately skewed and when sample size was large for MDU (e.g., N=100).

7) IMDU and EMDU were even worse than WMDU and should be avoided for Likert-type data. The reader should recall that the IMDU and EMDU procedures modeled the data in a completely different way than did the WMDU procedure.

CHAPTER VI

DISCUSSION

FA-PR vs. FA-PL

It was predicted that FA-PL should perform better than FA-PR. This prediction was made because previous literature had shown that FA-PL gave more accurate estimates of the pairwise true latent correlations and of the factor loadings than did FA-PR. When the recovered correlations were examined, Pearson correlations (rather than polychoric correlations) were indeed found to systematically underestimate the true latent correlations. It was also found that FA-PR rather than FA-PL systematically underestimated the true factor loadings. In addition, these underestimations were especially obvious when item responses were highly or differentially skewed. All of these findings were consistent with results of previous studies. Nonetheless, the current results showed that the systematic underestimation of both true latent correlations and factor loadings did not hamper the estimation of the true latent trait. Across almost all conditions set by the current study, FA-PR estimated the latent trait as accurately as did FA-PL. It seemed that the systematic underestimation of factor loadings was compensated by systematic overestimation of regression weights for factor scores when the latent trait was to be estimated. This compensation phenomenon was even found for dichotomous items in a later study (Chan, Hou & Chen, in preparation).

WMDU vs. IMDU vs. EMDU

As predicted, the WMDU model was found to outperform the two classical unfolding models in almost all experimental conditions. This result implied that, if procedures in the MDS family are to be employed for analyzing Likert-type data, the WMDU model should be a more reasonable alternative to the two classical unfolding models.

IMDU and EMDU were expected to completely fail in the situation where all item responses were normally distributed and to perform best when item responses were differentially skewed. This prediction was not confirmed. The performances of IMDU and EMDU were very unstable--they fluctuated from one condition to another. The underlying mechanism of the two classical unfolding procedures to model the monotonic relationships between the Likert-type items and the latent trait seems puzzling. The results of the current study suggested that IMDU and EMDU should not be applied to Likert-type items.

IRT vs. FA vs. MDS

Conceptually, IRT-GRM, FA-PL, FA-PR, and WMDU are all able to model both discrimination parameters and threshold values in Likert-type items. Their predicted relative merits were generally confirmed. IRT's non-linear formulation of the relationship between responses and latent traits seemed to have some advantages, especially when distributions of item responses were skewed. These advantages were perhaps due to the property that IRT item parameter estimates do not depend on the latent trait distribution of the calibration sample. The current study suggests that when distributions of item responses are moderately or highly skewed, IRT-GRM is the favored choice for estimating the latent trait.

When item responses are normally distributed, IRT-GRM, FA-PL, FA-PR, and WMDU are all reasonable choices, which may be decided according to other objectives.

SSI vs. Other Procedures

The common SSI practice for scaling the Likert-type items has usually been criticized for its strong assumption of equal intervals in the ordinal scales. The present study found that this strong assumption was not very harmful given that the number of response categories for each item was five. In the present study, the SSI procedure might be slightly worse than the IRT-GRM but was comparable with the two sophisticated FA procedures. The SSI procedure was even more robust against skewness than the two FA procedures. In addition, it was as good as or better than the WMDU and was definitely better than the two classical unfolding procedures. If computer time and computational simplicity were also taken into consideration, the SSI procedure became more attractive than the other six procedures.

Selection among Procedures

In addition to the dependent and independent variables explored by the current study, researchers may also need to consider the following factors when selection among the seven procedures is necessary:

a) *Model Features*. IRT, FA, and MDS have been developed by different schools in different statistical/psychometric traditions so that their objectives and features are also distinctive. IRT models have the well known property of "item-free person measurement" and "sample-free test calibration" (Wright, 1967). Therefore, IRT models are very suitable for dealing with test equating and adaptive testing. FA models are able to handle large numbers of latent variables and, if embedded in the context of structural equation modeling, are able to statistically test hypotheses about latent structure, i.e., the relationships among latent variables. Finally, MDS transforms psychological relationships to spatial relationships and can satisfy researchers who prefer pictures to numbers for revealing the meaning of data.

b) *Computer time (on IBM 3081)*. Generally speaking, the WMDU procedure was the most time-consuming procedure, IRT-GRM/FA-PL the second, FA-PR/IMDU/EMDU the third, and SSI the fourth. When the WMDU model did not fit the data well (e.g., in the situation where item responses were highly or differentially skewed), it could take as much as 800 seconds of CPU time given one analysis of 100 subjects by 24 items. The IRT-GRM and the FA-PL could require as much as 45 seconds of CPU time given 1000 cases, 24 items, and highly skewed responses.

c) *Computer's working memory*. Applying the WMDU model to Likert-type items, users will encounter numerous ties. The present study found that the amount of storage needed by the SAS PROC ALSCAL to cope with ties in 24 items by 100 cases is about 9000 bytes.

d) *Number of stimuli*. Most current computer programs for MDS have difficulty handling a data set with more than 200 stimuli.

Limitations

Regarding the latent space, the current study investigated only the unidimensional case. Some of the current findings may not be generalizable to multidimensional cases. For example, IRT-GRM which utilizes MML estimation based on the full information approach is preferable for long tests with few factors. MML is better with few factors because it requires integration over the entire factor space, which implies geometric increases in computation load as the number of factors increases. Multidimensional cases may be able to be investigated after IRT-GRM is developed to cope with multidimensional polychoric data.

Although Likert-type items are often based on 5-point response scales, general attitude measurements may also frequently be implemented with non 5-point scales. It is unknown whether or not the current findings would generalize to non 5-point scales. Although the current study had no reason opposing this kind of generalization, a further study is needed for providing some evidence. It is possible that, due to the increasing or decreasing number of parameters to be estimated, the relative merits of the seven statistical procedures may change with differing numbers of response categories.

There are many kinds of item-response distributions that are qualitatively different from one another. For example, U-shaped, uniform, or other irregular distributions are occasionally encountered in practice. The current study, however, investigated only normal and skewed distributions. It is still unknown how accurate the seven statistical procedures would be in estimating the latent trait from Likert-type data given those response distributions which were not investigated.

Invariance of the response threshold values across subjects is a major and neglected assumption of psychological research (Brady, 1989). It is also assumed by the current simulation processes. The effects of violating this assumption on the performances of the seven statistical procedures discussed above need to be explored in a further study. The current study predicted that, if this assumption is violated seriously, performances of all seven procedures except WMDU will be affected. WMDU may be appropriate for interpersonally incomparable data because in the "three way three mode" data set input for WMDU, the row stimuli which represent subjects/persons can be set to be row-conditional.

Although the current study tended to assume that the performance differences among investigated procedures were mainly due to mathematical modeling, estimation algorithms could be a minor confounding factor when comparisons were made across IRT, FA, and MDS. In the MULTILOG computer program for IRT-GRM, marginal ML was used to estimate item parameters while conditional ML was used to estimate person parameters. In the LISREL computer program for FA-PR and FA-PL, ML (or unweighted least squares if a correlation matrix was not positive definite) was used to estimate factor structures while least squares regression was used to estimate factor scores. In the ALSCAL computer program for IMDU, EMDU, and WMDU, alternating least squares was used to estimate both item coordinates and subjects' ideal points. Therefore, a portion of the performance differences across IRT, FA, and MDS might have been due to estimation rather than pure modeling differences. With a few samples, however, the author did compare the unweighted least squares solutions with ML solutions in FA but found little difference. Further studies are needed to isolate the effects of estimation from those of modeling.

REFERENCES

Anderson, T. W. (1959). Some scaling models and estimation procedures in the latent class model. In Probability and Statistics, The Harald Cramér Volume, ed. U. Grenander, Stockholm: Almqvist & Wicksell and Wiley, 9-38.

Andrich, D. (1978). A rating formulation for ordered response categories. Psychometrika, 43, 561-573.

Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. Applied Psychological Measurement, 12, 33-51.

Andrich, D. (1989). A probabilistic IRT model for unfolding preference data. Applied Psychological Measurement, 13, 193-216.

Babakus, E., Ferguson, C. E., Jr. & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. Journal of Marketing Research, 24, 222-228.

Bartholomew, D. J. (1987). Latent variable models and factor analysis. New York: Oxford University Press.

Bejar, I. I. (1983). Introduction to item response models and their assumptions. In R. K. Hambleton (Ed.), Applications of Item Response Theory. Vancouver, BC: Educational Research Institute of British Columbia.

Birnbaum, A. (1958). On the estimation of mental ability. Series Report No. 15. Project No. 7755-23. USAF School of Aviation Medicine, Randolph Air Force Base, TX.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick, Statistical theories of mental test scores. Reading, MA: Addision-Wesley.

Birnbaum, A. (1969). Statistical theory for logistic mental test models with a prior distribution of ability. Journal of Mathematical Psychology, 6, 258-276.

Bollen, K. A., & Barb, K. H. (1981). Pearson's r and coarsely categorized measures. American Sociological Review, 46, 232-239.

Brady, H. E. (1989). Factor and ideal point analysis for interpersonally incomparable data. Psychometrika, 54, 181-202.

Carroll, J. B. (1961). The nature of the data, or how to choose a correlation coefficient. Psychometrika, 26, 347-371.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrika, 35, 238-319.

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), Multidimensional scaling: Theory and applications for the behavioral sciences. Vol. I: Theory. New York: Seminar Press.

Chan, J. C., Hou, L. & Cheng, S. (in preparation). Pearson correlations vs. polychoric correlations: Options of factor analysis for ordinal data.

Cliff, N. (1989). Ordinal consistency and ordinal true scores. Psychometrika, 54, 75-91.

Clogg, C. C. (1979). Some latent structure models for the analysis of Likert-type data. Social Science Research, 8, 287-301.

Cohen, J. (1990). Things I have learned (so far). American Psychologist, 45, 1304-1312.

Coombs, C. H. (1964). A theory of data. New York: Wiley.

Coombs, C. H., & Smith, J. E. K. (1973). On the detection of structure in attitudes and developmental processes. Psychological Review, 80, 337-351.

Dancer, L. S. (1985). On the multidimensional structure of self-esteem: Facet analysis of Rosenberg's self-esteem scale. In D. Canter (Ed.), Facet Theory: Approaches to Social Research. New York: Springer-Verlag.

Davison, M. L. (1983). Multidimensional scaling. New York: John Wiley & Sons.

Davison, M. L. (1985). Multidimensional scaling versus components analysis of test intercorrelations. Psychological Bulletin, 97, 94-105.

Davison, M. L., & Srichantra, N. (1988). Acquiescence in components analysis and multidimensional scaling of self-rating items. Applied Psychological Measurement, 12, 339-351.

de Leeuw, J. (1983). Models and methods for the analysis of correlation coefficients. Journal of Econometrics, 22, 113-137.

Dodd, B. G. (1984). Attitude scaling: A comparison of the graded response and partial credit latent trait models. Doctoral Dissertation, University of Texas at Austin.

Dodd, B. G., Koch, W. R. & DeAyala, R. J. (1988). Computerized adaptive attitude measurement: A comparison of the graded response and rating scale models. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Everitt, B. S. (1984). An introduction to latent variable models. New York, NY: Chapman and Hall.

Fitzpatrick, S. (1989). Rasch model parameter estimation: A comparison of IRT and nonmetric multidimensional scaling methods. Paper presented in the annual conference of the American Educational Research Association.

Gillespie, M. K. (1989). A comparison of factor analysis and multidimensional scaling applied to survey data. Dissertation of the University of Texas at Austin, Austin, TX.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory. Boston: Kluwer.

Hildebrand, D. K., Laing, J. D., & Rosenthal, H. (1977). Analysis of ordinal data. Beverly Hills, CA: Sage.

Jenkins, G. D., Jr., & Taber, T. D. (1977). A Monte Carlo study of factors affecting three indices of composite scale reliability. Journal of Applied Psycholgoy, 62, 392-398.

Jöreskog, K. G., & Sörbom, D. (1979). Advances in factor analysis and structural equation models. Cambridge, MA: Abt Books.

Jöreskog, K. G., & Sörbom, D. (1984). LISREL VI: User's guide (3rd ed.). Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1988). PRELIS: A preprocessor for LISREL. Mooresville, IN: Scientific Software, Inc.

Kendall, M., & Stuart, A. (1977). The Advanced Theory of Statistics, vol. 1, 4th ed. New York: MacMillan.

Knol, D. L. & Berger, M. P. F. (1989). Empirical comparison between factor analysis and item response models. Paper presented in the annual conference of the American Educational Research Association.

Koch, W. R. (1983a). Likert-type scaling using the graded response latent trait model. Applied Psychological Measurement, 7, 15-32.

Koch, W. R. (1984). Degenerate solutions in multidimensional unfolding. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling. Psychometrika, 29, 1-27; 115-199.

Labovitz, S. (1967). Some observations on measurement and statistics. Social Forces, 46, 151-160.

Labovitz, S. (1970). The assignment of numbers to rank order categories. American Sociological Review, 35, 515-524.

Lawley, D. N., & Maxwell, A. E. (1971). Factor analysis as a statistical method. London: Butterworths.

Lehmann, D. R. (1974). Some alternatives to linear factor analysis for variable grouping applied to buyer behavior variables. Journal of Marketing Research, 11, 206-213.

Levin, J., Montag, I. & Comreg, A. L. (1983). Comparison of multitrait-multimethod, factor, and smallest space analysis on personality scale data. Psychological Reports, 53, 591-596.

Lingoes, J. C. (1972). A general survey of the Guttman-Lingoes nonmetric program series. In R. N. Shepard et al. (Eds.), Multidimensional scaling: Volume I (pp. 49-68). New York: Seminar Press.

Lissitz, R. W. & Green, S. B. (1975). Effect of the number of scale points on reliability: a Monte Carlo approach. Journal of Applied Psychology, 60, 10-13.

Loehlin, J. C. (1987). Latent variable models: An introduction to factor, path, and structural analysis. Hillsdale, NJ: Lawrence Erlbaum.

Lord, F. M. (1952). A theory of test scores. Psychometric Monograph, 7.

Lord, F. M. (1953a). An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. Psychometrika, 18, 57-75.

Lord, F. M. (1953b). The relation of test score to the trait underlying the test. Educational and Psychological Measurement, 13, 517-548.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading: Addison-Wesley.

MacCallum, R. C. (1974). Relations between factor analysis and multidimensional scaling. Psychological Bulletin, 81, 505-516.

Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Masters, G. N. (1988). Measurement models for ordered response categories. In R. Langeheine & J. Rost (Eds.), Latent Trait and Latent Class Models. New York: Plenum.

McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. Applied Psychological Measurement, 6, 379-396.

McDonald, R. P. (1985). Factor analysis and related methods. Hillsdale, NJ: Lawrence Erlbaum.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. Journal of Educational Statistics, 11, 3-31.

Mooijaart, A. (1985). Factor analysis for non-normal variables. Psychometrika, 50, 323-342.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. Journal of Econometrics, 22, 43-65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika, 49, 115-132.

O'Brien, R. M. (1979). The use of Pearson's r with ordinal data. American Sociological Review, 44, 851-857.

Olsson, U. (1979a). On the robustness of factor analysis against crude classification of the observations. Multivariate Behavioral Research, 14, 485-500.

Olsson, U. (1979b). Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika, 44, 443-460.

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. Psychometrika, 47, 337-347.

Pearson, K. (1913). On the measurement of the influence of 'broad categories' on correlation. Biometrika, 9, 116-139.

Ramsay, J. O. (1982). Some statistical approaches to multidimensional scaling data. Journal of the Royal Statistical Society, (A), 145, 285-312.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Denmark Paedogogiske Institute.

Rodgers, J. L., & Young, F. W. (1981). Successive unfolding of family preferences. Applied Psychological Measurement, 5, 51-62.

Roskam, E. E., & Lingoes, J. C. (1970). MINISSA-I: A FORTRAN IV (G) program for the smallest space analysis of square symmetric matrices. Behavioral Science, 15, 204-205.

Samejima, F. A. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika (Monograph Supplement No. 17).

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). Introduction to multidimensional scaling. New York: Academic Press.

Schlesinger, I. M., & Guttman, L. (1969). Smallest space analysis of intelligence and achievement tests. Psychological Bulletin, 71, 95-100.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance. I and II. Psychometrika, 27, 125-140, 219-246.

Silverstein, A. B. (1987). Multidimensional scaling vs. factor analysis of Wechsler's intelligence scales. Journal of Clinical Psychology, 43, 381-386.

Takane, Y. (1980a). Analysis of categorizing behavior by a quantification method. Behaviormetrika, 8, 75-86.

Takane, Y. (1980b). Maximum likelihood estimate in the generalized case of Thurstone's model of comparative judgment. Japanese Psychological Research, 22, 188-196.

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. Psychometrika, 52, 393-408.

Thissen, D. (1986). MULTILOG: A user's guide. Mooresville, IN: Scientific Software, Inc.

Thissen, D., & Steinberg, L. (1988). Data analysis using item response theory. Psychological Bulletin, 104, 385-395.

Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review, 34, 278-286.

Thurstone, L. L. (1931). Rank order as a psychophysical method. Journal of Experimental Psychology, 14, 182-201.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. Psychometrika, 17, 401-419.

Togerson, W. S. (1958). Theory and methods of scaling. New York: Wiley.

Wilson, D. T., Wood, R., & Gibbons, R. T. (1984). TESTFACT: Test scoring, item statistics, and item factor analysis. Mooresville, IN: Scientific Software.

Wright, B. D. (1967). Sample-free test calibration and person measurement. Proceedings of the 1967

Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service, 1986.

Young, F. W. (1970). Nonmetric multidimensional scaling: Recovery of metric information. _Psychometrika_, _35_, 455-474.

Young, F. W. (1984). The general Euclidean model. In H. G. Law et al. (Eds.), _Research methods for multimode data analysis_. New York: Praeger. pp. 440-465.

Young, F. W. (1987). _Multidimensional scaling: History, theory, and applications_. Hillsdale, NJ: Lawrence Erlbaum.

Young, F. W., & Lewyckyj, R. (1979). _ALSCAL-4: User's guide_. Carrboro, NC: Data Analysis and Theory Associates.

VITA

Chihyu Chan, the son of Chiu-Lin Chan and Chung-Mei Chan, was born April 4th, 1959 in Taipei, Taiwan, R. O. C. In 1978, he attended National Chengchi University, where he received a Bachelor of Education degree. From 1982 to 1984, he served in the army as a psychological counselor. From 1984 to 1986, he went back to the graduate school in National Chengchi University, where he received a Master of Education degree. His thesis was honored by the Shu Shiang-Yu Awards from the Chinese Psychological Association. In 1986 he obtained the Dr. Sun Yat-Sen Scholarship and began his doctoral study in the Department of Educational Psychology with a specialization in quantitative methods at The University of Texas at Austin, where he also served as a teaching assistant/statistical consultant.

Permanent Address: No. 234 Section 1 Chungshan Road
Shulin Town, Taipei County
Taiwan, R. O. C.

This dissertation was typed by the author.