KM 42 (1993)
pg.161-176

UNFOLDING MODELS FOR PICK ANY/N DATA: A SUMMARY AND DISCUSSION

Wijbrandt H. van Schuur[1][2]

Abstract

This paper first emphasizes the similarity between unfolding and cumulative scale analysis, and suggests some reasons for the difference in the popularity of the two models. It then summarizes the results presented by the authors of this issue for their unfolding analyses of the nuclear energy data and the car-use data, discusses similarities and differences among the findings, and suggests reasons for some of the differences found. Finally, it raises some more general issues concerning the data used, differences in parameter estimation, tests for model fit, and proposed generalizations.

Keywords: cumulative scaling, method comparison, parallelogram analysis, pick any/n data, unfolding.

[1] Requests for reprints should be sent to: Wijbrandt H. van Schuur, Department of Sociology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands. Tel 31-50-636463-636226(FAX), email: WSCHUUR@RUG.NL.

[2] I would like to thank Melissa Bowerman for her editorial help.

1. Introduction

1.1 The Relationship between Unfolding and Cumulative Scaling

The deterministic unfolding model for dichotomous data developed by Coombs (1964) and the deterministic cumulative scaling model developed by Guttman (1950) are rather similar. Both models require variables with two response categories, the 'positive' response (generally denoted as '1') and the 'negative' response (generally denoted as '0'). (Typical positive responses in the behavioral sciences include 'gives a correct answer to a question about facts', 'agrees with an evaluative statement', 'has performed a certain action', and 'has voted in favor of a certain measure'; negative responses are the converse of these). In both models it is assumed that the positive response to one variable from a set of questions about facts, statements, actions, or measures is related to the positive response to other variables from the same set. And in both models the variables can be ordered in such a way that the positive response to one variable implies the positive response to an adjacent variable, as illustrated in Table 1.

Table 1
Data Sets that Perfectly Fit the Cumulative Scaling Model and the Unfolding Model for Dichotomous Data

Subjects	Cumulative Scale A B C D E	Unfolding Scale A B C D E
1	0 0 0 0 0	1 1 0 0 0
2	1 0 0 0 0	0 1 1 0 0
3	1 1 0 0 0	0 1 1 1 0
4	1 1 1 0 0	0 0 1 1 0
5	1 1 1 1 0	0 0 1 1 1
6	1 1 1 1 1	0 0 0 1 1

The difference between the two models is that in cumulative scaling the positive response to a variable is interpreted as a *dominance* relation between subject and items (i.e., a subject who gives the positive response to an item 'dominates' this and all 'easier' items), whereas in unfolding the positive response is interpreted as a *proximity* relation between subject and items (i.e., a subject who gives the positive response to an item also gives the positive response to adjacent items, but not to items that are more distant in either direction). The scale score of subjects in the cumulative model is defined as a function of the sum score of their responses, whereas the scale score of subjects in the unfolding model depends on the specific items to which they give the positive responses.

1.2 Unfolding Models as Derived from Cumulative Scaling Models

Almost all the models discussed in this issue are related to developments in cumulative scaling. Formann's quasi-deterministic latent class model shows the influence of Goodman's latent distance model (Goodman 1975), Croon's ordinal latent class model is related to his own earlier model for cumulative data (Croon 1991), Böckenholt's parametric unfolding latent class model is similar to Rost's (1988) model, Van Schuur's and Post & Snijders's ordinal latent trait models are related to Mokken's (1971) model, and Hoijtink's and Verhelst & Verstralen's parametric latent trait models show the influence of the Rasch model (Rasch 1960) and the Partial Credit Model (Masters 1982), respectively. The only exception is Van Blokland-Vogelesang's model, which is based on full rank order data.

The authors of this issue differ on whether both unfolding models and cumulative models should be considered Item Response Theory (IRT) models. For example, Verhelst & Verstralen suggest that the term IRT should be reserved for the (cumulative) Rasch model, whereas Hoijtink argues that since the unfolding (or parallelogram) model also links items to responses, it should also be called an IRT model. Andrich (1988) proposed a compromise: calling parallelogram models 'Person-Item-Response-Theory (PIRT) models'. But this is not really satisfactory, since each IRT model includes persons as well as items, and so could equally well be called a PIRT-model.

Both the authors of the latent class analysis models and Post & Snijders regard their models as the offspring of latent structure analysis, developed by Lazarsfeld and Henry (1968). I agree with attempts to unify terminology. Efforts to make very fine distinctions (e.g., refusing to consider parallelogram analysis as a form of unfolding, as in Cliff et al. 1988) lead to confusion among researchers trying to decide which model to apply to their data.

1.3 Why Unfolding Never Became as Popular as Cumulative Scaling

The cumulative scaling model became known as "the Guttman scale". It developed into a number of probabilistic models, of which the parametric Rasch scale (Rasch 1960) and the nonparametric Mokken scale (Mokken 1971) are perhaps the best known. In contrast, the unfolding model -- or the parallelogram analysis model, named after the parallelogram shape of '1's in the data matrix -- never became known as "the Coombs scale". Programs to perform Guttman Scaling began to appear in statistical packages, but programs to perform parallelogram analysis did not. User-friendly software for parallelogram analysis did not become widely available before the 1980's.

One reason the unfolding model did not become as popular as the cumulative model is probably that in many of the data sets investigated, the relationship between the dichotomous variables was better captured by reference to dominance than to proximity. This is especially true in the area of ability testing. The unfolding model has also perhaps developed more slowly than the cumulative model because Coombs initially formulated it for full rank order data rather than dichotomous data. Rank order data are not very common to begin with, and, for a long time starting in the sixties, such data could only be unfolded as a byproduct of multidimensional scaling; representations were often degenerate. The unfolding of rank order data with many ties (one way to describe dichotomous data) was therefore even less popular. The model developed by Van Blokland-Vogelesang for full rank order data is an improvement over these previous MDS models.

The prevailing opinion among applied researchers still seems to be that full rank orders of data should be unfolded whereas dichotomous data should be cumulatively scaled. Researchers further assume that Likert-type rating scales should be factor analyzed, rather than cumulatively scaled or unfolded. In fact, however, dichotomous data or Likert-type data can and often should be unfolded. The unwarranted insistence that the form in which data are collected determines the way in which they should be analyzed has done damage to substantive theory formation. Even though as early as 1960 Coombs and Kao warned against using factor analysis for data that conform to the unfolding model, this warning is rarely heeded. Coombs and Kao claimed (and Ross and Cliff, 1964, showed for a specific case) that the factor analysis of unfoldable data leads to an extra factor (or dimension) -- one more than is needed in an unfolding representation. When the unfolding representation is unidimensional, factor analysis thus results in two apparently independent factors. Ignorance of the "extra factor" phenomenon has led researchers to puzzle over the seeming independence of such bipolar concepts as positive and negative mood states, liberalism and conservatism, work intrinsic and work extrinsic work satisfaction, and masculinity and femininity, to mention only a few. Given this problem it is important for researchers to realize that unidimensional unfolding models are now available that do not require full rank orders of data, and, more generally, that the form in which data are collected is in principle independent of the way they can be analyzed.

2. The Nuclear Energy Data

The nuclear energy data consist of responses to five questions selected from a questionnaire that measured the general attitudes toward nuclear energy of 600 Austrian respondents (Formann 1988).

2.1 Scale Values for the Items and Interpretation of the Scale

All the models agree about the appropriateness of the unfolding model to the nuclear energy data, and about the order in which the items form an unfolding scale (see Table 1 in Hoijtink's editorial for the order). The scale is unequivocally interpreted as measuring opinions about nuclear energy ranging from strongly in favor (item A) to strongly opposed (item E).

2.2 Goodness-of-fit of Individual Items

The models of Hoijtink, Van Schuur, Post & Snijders, and Verhelst & Verstralen explicitly analyze the fit of individual items to the model. Hoijtink picks out items B or E as fitting the model poorly. Van Schuur points only to item B. Post & Snijders recommend dropping either item B or item C, whereas Verhelst & Verstralen identify D or E as a poorly-fitting item. Most of the other models give some information that might be used to corroborate information about individual items, although this remains implicit in their contributions. For instance, the results of Böckenholt's and Croon's unrestricted latent class analysis are similar -- see Table 2 for the probabilities of positive responses to the items by subjects in each latent class in the two analyses. Rowwise the probabilities do indeed conform to the expected pattern of single-peaked monotonicity. Columnwise, however, item B destroys the single-peaked monotonicity pattern in both class I and class IV. Van Blokland-Vogelesang's procedure allows us to compare the best- fitting four-item scale with the best-fitting five-item scale. On inspecting the data (as given by Hoijtink in Table 1), we find that among the 600 response patterns, there are 114 patterns (=19%) that violate the deterministic unfolding model. If item B is removed, only 18 of the 600 patterns (3%) contain a model violation. This is less than would be achieved by removing any other item. Van Blokland-Vogelesang's procedure thus identifies the best fitting four-item scale as ACDE; that is, a scale without item B.

Table 2
Percentage of Subjects in Each Class who Give the Positive Response to Each of the Five Items

	BÖCKENHOLT Classnrs				CROON Classnrs			
Itemnr	I	II	III	IV	I	II	III	IV
A	54	46	8	0	54	46	8	0
B	49	62	28	15	49	63	28	15
C	77	97	73	0	78	97	72	0
D	27	91	100	14	20	91	100	14
E	0	26	98	100	0	27	98	100

Formann uses two criteria to assess individual items. First, the area along the latent dimension in which the positive response to each item is given should be neither too small (as for item A) nor too large (as for item C). This is not a criterion of fit, but rather a criterion to insure that an item is useful in discriminating among different latent classes of subjects. Second, the probability of a positive response to all items should be substantially higher than the probability of a negative response. On the basis of this criterion Formann singles out item B as a poorly fitting item.

In conclusion, most of the models agree on the bad fit of item B. But they differ in how they define bad fit. For Hoijtink, the fit of item B is bad because its item characteristic curve is too flat (that is, subjects with different values on the latent dimension give the positive response with similar probabilities). Table 2 corroborates this finding to some extent: the four classes do not differ much in the percentages of positive responses for item B. Post & Snijders' Table 4 also shows the flat ICC of item B.

Hoijtink found that item E had a flat ICC, as well as item B. Verhelst & Verstralen agree with Hoijtink on the poor fit of item E, but none of the other authors whose analyses corroborated the flatness of item B's ICC suggest that item E behaves similarly. So we are still left with the question whether the poor fit of item E may be due to violations of other assumptions in the parametric models. For instance, how sensitive are the parametric models to any kind of deviation from the expected shape of the ICC of an item?

Van Blokland-Vogelesang and Van Schuur single out item B as badly fitting on grounds that many subjects gave a positive response to item A on the left and to items C, D, or E on the right, but not to B in the middle. None of the other items in the final scale were weak according to this or any of their other criteria of fit.

2.3 Distribution of Subject Scale Values

How are the subjects distributed along the scale? The latent class analysis models differ widely in the number of ordered latent classes (i.e., the number of different scale values for subjects) needed to represent the data. Croon uses as scale values the order of the four classes that fit the unrestricted model, and Böckenholt specifies the subject parameters further by the unfolding threshold model (UT) and the unfolding power model (UP), in which cases a representation in only two latent classes (i.e., with only two homogeneous groups of subjects) suffices.

Formann identifies nine latent classes, one of which contains unscalable subjects. He finds many unscalable response patterns (42%). This is in part due to one of the assumptions of his model that in effect restricts the range of positive responses each subject can give, rendering certain patterns (00000, 01000, 00100, 00010, 00110, 01111, 11110, and 11111) unscalable even though they do not actually violate a deterministic unfolding model.

Van Blokland-Vogelesang and Van Schuur do not specify subjects' scale values in their papers, although both have procedures for calculating them. Van Blokland-Vogelesang's procedure specifies at most $5(5-1)/2 + 1 = 11$ different scales values for subjects with five items. In her procedure, subjects with imperfect patterns are assigned the scale value of the perfect pattern that can be achieved with the minimum number of inversions. For dichotomous data there are often several ways to turn an imperfect pattern into a perfect one (e.g., 11101 can be converted equally easily to 11111 or 11100; these two of course have different scale values). In such cases the less extreme pattern is apparently chosen. When I analyzed the nuclear energy data with Van Blokland-Vogelesang's UNFOLD program, I found that, as a consequence of this procedure, 47% of subjects were assigned the same scale value.

Van Schuur's procedure specifies at most $2*5 - 1 = 9$ different scale values for an unfolding scale with five dichotomous items. Each imperfect pattern receives a unique scale value, so there is a wider distribution of scale values of subjects than in Van Blokland-Vogelesang's model. Still, more than 80% of all subjects are assigned to only four classes (scale scores 4 (13%), 5 (25%) 6 20%) and 7 (24%)).

Hoijtink does not specify the scale scores for the subjects (his EAP-scores), but gives a discrete estimate of the density function. Two of the 7 nodes carry 93% of the subject density, which, like the results of the previous models, suggests a very high concentration of subjects in the center of the scale.

In conclusion, although the models under consideration differ in their assignment of scale values to subjects, they agree that subjects may be described in terms of a small number of values. Latent class analysis (i.e., the models of Formann, Croon and Böckenholt) and latent trait analysis (i.e., the models of Van Blokland-Vogelesang, Van Schuur, Post & Snijders, Hoijtink, and Verhelst & Verstralen) differ fundamentally, however, in their approach to the optimal number of different scale values of subjects. In an ideal latent class analysis the number of classes is much smaller than in the ideal latent trait analysis for the same number of items. In latent class analysis, researchers look for as much homogeneity among subjects as possible, and so try to represent subjects in a small number of classes. In latent trait analysis, in contrast, researchers try to discriminate between subjects to prevent local stochastic independence from becoming global independence.

3. The Car-use Data

The car-use data consist of two independent data sets, each comprising 300 Dutch subjects who, either before or after a governmental information campaign, were asked to give their opinion on the use of cars in the context of protection of the environment (Doosje & Siero, 1991). One goal of the research was to determine whether the campaign had been successful.

3.1 Scale Values for the Items and Interpretation of the Scale

Most authors agree with the order of the items along the unfolding scale given by Hoijtink in his editorial (Table 4), except that they would reverse items 1 and 2 and items 7 and 8. The scale is unequivocally interpreted as measuring opinions ranging from strongly in favor of measures against car use to strongly opposed.

The latent class models have problems in reaching this solution. They need to assign probabilities to each possible response pattern in each latent class. This becomes difficult when the number of items is 10 or more, with the number of possible response pattern thus at least 1024. If the number of actually observed response patterns is small, the models cannot function properly. Böckenholt therefore decides to use only six of the ten items; his model orders them in agreement with the results mentioned above. Croon does not explicitly discuss the ordering of the items. However, in addition to insisting on an ordering of the choice probabilities of each item across the classes, he could also have looked at a single-peaked ordering of the choice probabilities of the items within each class. The order given above largely satisfies the requirement that there be a single-peaked function to the order of the probabilities of the positive response to the items within each class. Following the same strategy as Böckenholt, Formann searches for the best maximal subset of the items, and ends up with a scale of five items. Like Croon, he does not discuss whether the probabilities of the positive response to the items form a single-peaked pattern within each class. However, his Table 4 shows that when the five items are ordered according to the results of most other authors, the single-peaked pattern is violated several times.

Four authors formally compare the unfolding scales they identify for the pre- and post-information campaign data. Formann and Böckenholt demonstrate the use of simultaneous latent class analysis to compare the different groups. Hoijtink and Verhelst & Verstralen both use a likelihood ratio test to check whether the model parameters are the same for both data sets; they come to different conclusions, however. Hoijtink concludes that the item- and γ-parameters are the same in the pre- and post-campaign data, but that the subject distribution is different. In contrast, Verhelst & Verstralen conclude that the model parameters differ, but the subject distribution is almost the same. From a substantive perspective Hoijtink's approach seems to be easier to interpret than that of Verhelst & Verstralen.

The four authors who use nonparametric models have no formal test with which to compare the pre- and post-campaign results. They must therefore resort to the well-known 'eyeball technique', which leads them to conclude that the pre- and post-campaign data sets do not differ substantially in their representation of the items.

3.2 Goodness-of-fit of Individual Items

What do the different models say about the fit of the individual items? Hoijtink is alone in arguing explicitly that item 1 cannot be represented. He rejects it on grounds that its item characteristic curve (ICC) is too flat. Böckenholt's Tables 6 and 7, Croon's Table 4, Van Schuur's Table 3, and Post & Snijders' Table 8 also show a flat curve for item 1. However, item 1 is not the only item with a relative flat ICC: on the basis of the corroborating evidence for item 1 from the tables by the other authors, we might also single out item 2 as having a flat ICC according to the tables. In the ordinal models of Croon and Van Schuur, the specific shape of the ICC of item 1 is not a problem as long as it does not disturb the order of the probabilities. If both 'pro- car' and 'pro-environment' subjects agree with this item, its only possible position on the scale is in the middle, which is indeed where the other models also represent it.

3.3 Distribution of Subject Scale Values: was the Campaign Effective?

If the information campaign was successful, the mean scale value of the post-campaign data should be lower than that of the pre-campaign data. Van Schuur and Verhelst & Verstralen find that although the difference in the means is in the expected direction, it is not statistically significant. According to Böckenholt, Hoijtink, and Formann, however, the difference is significant (just barely for Formann). When I ran Van Blokland-Vogelesang's UNFOLD program on the eight items that form an identical scale for both the pre- and post campaign data, I also found a large difference in scale values in the expected direction (from 56.1 to 43.6; $t=-8.92$, $p < 0.001$). According to Van Blokland-Vogelesang, the standard deviation of the distribution of subjects' scale values is smaller in the post- than the pre- campaign data set. This finding -- also mentioned by Böckenholt and Verhelst & Verstralen and inferable from Hoijtink's Table 7 -- suggests that another result of the information campaign may have been greater agreement in the public about the effects of car use on the environment.

Different models thus lead to different conclusions about whether the campaign had any effect. Such disagreement about scale value assignment is not unprecedented. For example, Van Schuur (in press) correlated the scale values of analyses of an 'androgyny' data set carried out with the models of Hoijtink, Van Schuur, and Van Blokland-Vogelesang. He found that although the scale values of the first two models correlated highly with each other (0.93), they correlated only moderately with the scale values of the third model (about 0.7). Further study is clearly needed to identify the sources of discrepancy. For example, attention should be paid to the number of items picked in the response patterns compared to the total number of items, the degree of fit of the models, the assignment of scale values to (deterministically) imperfect response patterns, and the number of items in the scale.

3.4 The Relationship Between Unfolding and Cumulative Scaling Revisited

A number of the authors (Croon, Formann, Van Schuur, Post & Snijders) have observed that although the car-use data can be interpreted as one unfolding scale, it can also be interpreted as two cumulative scales glued together: one ranging from the least popular item (i.e., the item least likely to receive a positive response) on the left to the most popular item in the middle of the unfolding scale, and the other ranging from the next most popular item in the middle of the unfolding scale to the least popular item on the right. Which of the two accounts is preferable? This question is especially important if we want to regard the scale values of subjects as values of a new variable that is to be used in further analyses, since scale values are interpreted differently under the two accounts. I return to this question shortly. But first let us consider how we can determine whether a given unfolding scale is also open to interpretation as two cumulative scales.

Some of the diagnostics proposed by Post & Snijders and by Van Schuur are relevant to this decision. An unfolding scale can also be interpreted as two cumulative scales if 1) the correlation matrix of the two sets of cumulative items includes a block of negative correlations that would become positive after the appropriate recoding of one of the item sets; 2) the 'matrix of p-values' based on the conditional adjacency matrix shows the highest p-values only for the left-most and right-most items; and 3) the dominance matrix does not show a pattern of characteristic monotonicity around the diagonal, but rather a pattern of monotonicity going from the left-most item up to the most popular item, and from that item down to the right-most item.

About half the unfoldable 'pick any/n' data sets that I have analyzed satisfy these diagnostics and so could be interpreted either as a simple unfolding scale or as two cumulative scales. The problem of deciding between these analyses is thus not trivial.[3] Are there advantages to one account or the other? On grounds of parsimony, the unfolding account is to be preferred. But the cumulative scale account has an advantage as well. In a deterministic cumulative scale, subjects are represented between the last item to which they give the positive response and the first item to which they give the negative response. If two cumulative scales are regarded as glued together, subjects will receive two scale values: one for the left-hand scale and one for the right-hand scale. These values in effect delimit the *range* of items to which the subject gives the positive response.

In an unfolding scale no range information is specified: subjects are simply represented 'in the middle of' the items to which they give the positive response. But in many circumstances investigators might find it useful to know the range of items to which individual subjects give their positive responses.[4] A range parameter that differentiates subjects with small ranges from subjects with larger ranges might also lend itself to interesting interpretations; for example, the size of the range may depend upon variables such as 'strength of feeling of interest', 'education or knowledge', or 'dogmatism or pragmatism'. Niemöller (1992) recently showed that a range parameter could be useful in explaining voting behavior on the basis of positions of subjects and political parties along a left-right dimension.

To further develop the theory that links persons' responses to items, it is imperative to understand both the unfolding and the cumulative response mechanism in a joint framework. It may ultimately be possible to combine the advan-tages of a range specification with the parsimony of an unfolding scale.

---

[3] An alternative approach to this problem might be to reverse the coding of one of the two sets of cumulative items and regard them all as members of the same cumulative scale. But available analyses (e.g., Kruijtbosch, 1992) show that it is by no means self-evident that the two sets do indeed form one cumulative scale after the codes of one set have been reversed. More work on this issue is needed.

[4] The only unfolding model that at present includes a subject-range parameter is DeSarbo and Hofmann's (1986); as far as I know, however, an algorithm for the model only exists in experimental form.

4. General Issues: Some Brief Remarks

4.1 The Data Used in Tests of Model Fit

In a Rasch analysis the 'pick 0/n' and 'pick n/n' response patterns are not included in tests of model fit. In a Mokken analysis an overrepresentation of such patterns inflates the fit of the model. It can be argued that, for the same reasons as in these models, 'pick 0/n' and 'pick n/n' patterns, as well as the 'pick 1/n' patterns, should also be discarded from a test of model fit in unfolding analyses. But most models do not report precautions of this kind.

4.2 Parameter Estimation

In most unfolding procedures, item parameters are estimated first and subject parameters are estimated only later, if at all. Like the cumulative IRT-models, however, unfolding can be regarded as an important intermediate analytic step that provides the researcher with a reliable measurement of latent variables that cannot otherwise be obtained. The emphasis on representing and interpreting the items rather than the subjects may be related to the MDS-tradition in which finding the interpretation of the unfolding dimensions has traditionally been more important than using the subject scale values in subsequent research.

In cumulative scaling procedures, scale values for 'pick 0/n' and 'pick n/n' patterns can be defined (even though these patterns are not included in tests of model fit). In unfolding procedures, the 'pick 1/n' and 'pick n/n' patterns can also be given a scale value, but the 'pick 0/n' pattern remains ambiguous. The models discussed differ in whether they assign an extreme (positive or negative) value to this pattern or interpret it as missing datum.

4.3 Model Fit

If a data set cannot be represented in full by a unidimensional unfolding model, several options are often open to researchers. For instance, they can use a model with more parameters, a weaker model, a model for a subset of the items, or a model for a subset of the subjects. Most of the models discussed emphasize the first three strategies. The proposals for item fit, especially by Post & Snijders, may be fruitfully implemented in some of the other models as well. Only Formann is explicit about the possibility that some *subjects* should be considered unscalable. The choice of whether to attribute bad model fit to items or to subjects is an important strategic choice. Often researchers do not want to sacrifice subjects since this may damage the properties of their sample. Items, on the other hand, are generally not sampled from a population of items, but are selected as the most reliable ones for measuring the latent trait. If they are not useful for this purpose, they can be eliminated with little loss.

Person fit can be operationalized readily in the latent class procedures. Subjects are assigned the scale value of the class for which the probability of their response pattern is highest. If a subject's response pattern is not clearly more probable for one class than another, the subject fits the model poorly.

Van Blokland-Vogelesang's and Van Schuur's procedures can easily be extended to include a measure of person fit in terms of the amount of model violation in each response pattern. The models of Hoijtink and Verhelst & Verstralen can probably also be extended, following Hoijtink's suggestions for person fit in the Rasch model (Molenaar and Hoijtink, 1991).

4.4 Additional Approaches

Croon discusses the possibility of using 'pick k/n' data. Models that can be applied to such data are useful because they allow researchers to unfold full or partial rank ordered data by dichotomizing them into the k most preferred and n-k least preferred items. This option is included in some of the models discussed (e.g., those of Van Blokland-Vogelesang and Van Schuur); the latent class models can also easily be extended. A further extension would be to generalize 'pick k/n' models to 'rank k/n' models on the basis of the same principles.

Some models can be extended for use with data involving ratings. All latent class models can be used for ordered multicategory data, as can Van Blokland-Vogelesang's latent trait model. Van Schuur's latent trait model has recently been generalized to multicategory data (Van Schuur in press). Presumably Verhelst & Verstralen's model would also allow this extension.

5. Conclusion

In this issue of Kwantitatieve Methoden eight different approaches to the unfolding analysis of 'pick any/n' data have been presented. A number of additional approaches have been omitted due to lack of space; the most notable omissions are the models of Andrich (1988), Andrich and Guanzhong (1991), Brady (1990), Cliff et al. (1988), Davison (1980), DeSarbo and Hoffman (1986), and Heiser (1981) (see De Soete et al., 1989, for some additional models). However, all the present authors have presented their own approaches in the broader context of the current literature. This issue therefore gives a useful overview of ongoing work in unfolding analysis.

One may wonder whether it is fair to compare the models on the basis of their performance on only two data sets. What would have happened, for instance, with a data set in which 12 of 17 items form an unfolding scale but the other five do not? Or with a data set consisting of 4 reasonably unfoldable items and only 20 subjects?

This issue, and particularly this last chapter, should not be read as a consumer guide. As one of the eight contributors of a model I am not in a position to write such a guide. In comparing the results of the models I have proposed some criteria that might be used in evaluating them. If we can agree on such criteria, the next step is to compare models by applying them to synthetic data sets in which the properties of the data we want to investigate can be controlled.

The unfolding models presented here form a welcomed addition to the methodological toolkit of the behavioral scientist. More work is in progress; an important role in this development has been played by the working group on 'Preference Analysis' of the SoMO (the methodological branch of the sociological unit (SSCW) of the Dutch Science Foundation (NWO) under the leadership of Van Blokland-Vogelesang. I am looking forward to seeing the new unfolding models applied increasingly frequently to types of data for which unfolding analysis has often been overlooked in the past.

References

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. & Guanzhong, L. (1991). A latent trait model for unfolding based on a Rasch model for ordered response categories. Manuscript submitted for publication.

Brady, H.E. (1990). Traits versus issues: factors versus ideal-point analysis of candidate thermometer ratings. Political Analysis, 2, 97-129.

Cliff, N., Collins, L.M., Zatkin, J., Gallipeau, D., and McCormick, D.J. (1988). An ordinal scaling method for questionnaire and other ordinal data. Applied Psychological Measurement, 12, 83-97.

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

Coombs, C.H. and Kao, R.C. (1960). On a connection between factor analysis and multidimensional unfolding. Psychometrika, 25, 219-231.

Croon, M.A. (1991). Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. British Journal of Mathematical and

Statistical Psychology, 44, 315-331.
Davison, M.L. (1977). On a metric, unidimensional unfolding model for attitudinal and developmental data. Psychometrika, 42, 523-548.
Davison, M.L. (1980). A psychological scaling model for testing order hypotheses. British Journal of Mathematical and Statistical Psychology, 33, 123-141.
DeSarbo, W.A. & Hoffman, D.L. (1986). The simple and weighted thresholds model for the spatial representation of binary choice data. Applied Psychological Measurement, 10, 247-264.
De Soete, G., Feger, H., and Klauer, K.C. (eds.) (1989). New developments in psychological choice modelling. Amsterdam: North Holland.
Formann, A.K. (1988). Latent class models for non-monotone dichotomous items. Psychometrika, 53, 45-62.
Goodman, L.A. (1975). A new model for scaling response patterns: an application of the quasi-independence concept. Journal of the American Statistical Association, 70, 755-768.
Guttman, L. (1950). The basis for scalogram analysis. In: S.A. Stouffer, (Ed.), Measurement and prediction. New York: Wiley, 216-257.
Heiser, W.J. (1981). Unfolding analysis of proximity data. Leiden: University of Leiden.
Kruijtbosch, M. (1992). Het ontvouwingsmodel als alternatief voor Bradburn's Affect Balans Schaal, m.b.t. het meten van 'mood states' in tien landen van Europa. Groningen: doctoraalscriptie vakgroep Sociologie.
Lazarsfeld, P.F. and Henry, N.W. (1968). Latent structure analysis. Boston: Houghton Mifflin.
Masters, G.N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.
Mokken, R.J. (1971). A theory and procedure of scale analysis. New York /Berlin: De Gruyter (Mouton).
Molenaar, I. & Hoijtink, H. (1990). The many null distributions of person fit indices. Psychometrika, 55, 75-106.
Niemöller, B. (1992). Left and right in wider perspective. Acta Politica, 27, 333-345.
Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish institute for educational research.
Ross, J. and Cliff, N. (1964). A generalization of the interpoint distance model. Psychometrika, 29, 167-176.
Rost, J. (1988). Rating scale analysis with latent class models. Psychometrika, 53, 327-348.
Van Schuur, W.H. (in press). A comparison of unidimensional unfolding models: masculinity and femininity or androgyny? In: J.A.J. Oud & R. Van Blokland-Vogelesang (Eds.), Proceedings of the SMABS 1992 conference in Nijmegen.
Van Schuur, W.H. (in press). Nonparametric unidimensional unfolding for multicategory data. Political Analysis, 4, 1993.