Nonparametric Unidimensional Unfolding for Multicategory Data

Wijbrandt H. van Schuur

Abstract

This article describes a nonparametric unidimensional unfolding model for dichotomous data (van Schuur 1984) and shows how it can be extended to multicategory data such as Likert-type rating data. This extension is analogous to Molenaar's (1982) application of Mokken's (1970) nonparametric unidimensional cumulative scaling model. The model is illustrated with an analysis of five-point preference ratings given in 1980 to five political presidential candidates by Democratic and Republican party activists in Missouri.

. . . It has something to do with numbers . . .
—"The Politics of Goofing Off," a review of the 1991 APSA Meeting
Washington Post, August 31, 1991

Introduction

Many variables that social science researchers want to analyze cannot be measured with a simple device, like a reading on a dial or an answer to a question. The researcher who wants a valid and reliable measurement—for example, of an individual's ideological position or amount of political knowledge—needs to proceed indirectly. For measuring social science vari-

I would like to thank Melissa Bowerman for her editorial help, Ivo Molenaar, Frans Stokman, and two anonymous reviewers for their comments on an earlier version of this article, and Bas van Rens for writing additional software. The data used in this article are available either from the editor of *Political Analysis* or from the author.

ables, the researcher's methodological tool kit contains such time-honored techniques as reliability analysis, principal component analysis, and factor analysis; these are included in the major statistical packages. This article deals with some new techniques that extend the tool kit.

The purpose of reliability analysis, component analysis, and factor analysis is to determine the homogeneity of a number of manifest variables, which are often answers to questions in a questionnaire and are generally known as "items." This is done by looking at their intercorrelations. If a set of items is sufficiently homogeneous, we can take the (weighted) sum of responses to the items as the measurement value on a new variable, which is called a latent variable. Examples of latent variables include "political knowledge" and "ideological position." Political knowledge can be measured by administering a number of questions about political facts (e.g., the names of politicians). The (weighted) number of correct answers is taken as a measure of political knowledge, with a low value meaning little political knowledge and a high value meaning a great deal. Ideological position can be measured similarly with questions about preferences for various politicians. Answers are coded as "high" or "low." The extent of preference for conservative over liberal politicians is a measure of the position of a respondent's position on a liberal-conservative scale, with a low value corresponding to a liberal position and a high value a conservative position.

In the last decades, many popular procedures for item analysis have come under increasing criticism. Psychologists and educational researchers who specialize in test theory[1] have criticized them for their strong assumptions (e.g., about level of measurement and about subject distributions on the manifest variables), and they have introduced new proposals, generally known as Item-Response Theory (or IRT) models. IRT models can be classified on the basis of two criteria: (1) whether the probability of a positive response to the manifest variables is related to a subject's value on the latent variable in a monotonic (e.g., linear) or single-peaked (e.g., quadratic) way; and (2) the measurement level of the latent variable (ordinal or interval). I will briefly explain these criteria and show how the model I propose fits into the IRT tradition.

The first criterion distinguishes the cumulative model from the unfolding model. Items indicating political knowledge are assumed to follow the cumulative model; that is, a positive (i.e., "correct") response to a question about a political fact is expected to occur with increasing probability for subjects with more political knowledge. The classical models mentioned previously (e.g., reliability analysis) are all cumulative models, as is the Guttman scale.

1. One of the best examples is Fischer 1974. There is probably still a market for an English translation of this book, which is written in German.

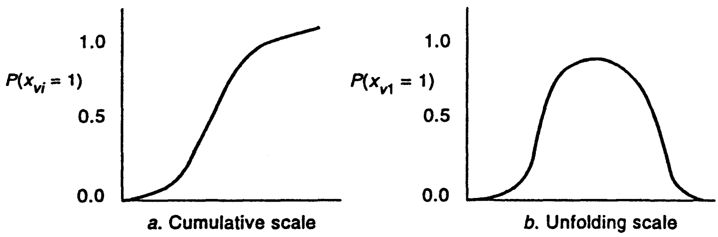

Fig. 1. Probability of positive response to item *i* by subjects with different scale values on the latent continuum
***a.* Cumulative scale**
***b.* Unfolding scale**

Items indicating ideological positions follow the unfolding model. The probability of a positive (i.e., "high") answer to such items does not increase monotonically for subjects with increasing values on a liberal-conservative scale but, rather, shows a single-peaked function. For instance, liking for an independent candidate for president of the United States (e.g., Anderson in 1980) will be low among those who identify strongly with the Democrats, higher for subjects who are "between" the Democrats and the Republicans, and low again for those who identify strongly with the Republicans. Figures 1a and 1b show the item-characteristic curves (the probability of a positive—i.e., "correct" or "high"—response for subjects with different values on a latent variable) for the cumulative model and the unfolding model.

The second criterion distinguishes models in which subjects are measured on an interval—versus an ordinal—scaled latent variable. Interval-scaled latent variables produce numerical scale scores for subjects and items according to the parametric model that is specified. Ordinal-scaled latent variables, in contrast, produce only rank orders of scale values for subjects and items, following the nonparametric model specification (i.e., the nonparametric relationship between the manifest variables and the latent variable). IRT models for cumulative scaling with an ordinal latent variable include the deterministic Guttman scale (Guttman 1950) and its nonparametric probabilistic analogue, the Mokken scale (Mokken 1970). Ordinal IRT models for unfolding include Coombs's deterministic parallellogram model (Coombs 1964) and the MUDFOLD model (van Schuur 1984). IRT models in which the latent variable is measured at the interval level include the Rasch model (Rasch 1960) for cumulative scaling and the models of Poole (1984), DeSarbo and Hoffman (1986), Andrich (1988), Hoijtink (1990), and Brady (1990) for unfolding. Models for ordered latent class analysis have recently been developed in which the ordering of subjects (the classes) depends on the cumulative model (e.g., Croon 1990; Rost 1988) or on the unfolding model (e.g., Böckenholt and

Böckenholt 1990; Formann 1988). Software is available for almost all of these models, often in user-friendly and well-documented form.[2]

Researchers prefer to measure their latent variables at the highest possible level of measurement, that is, with a parametric model. But the simpler, nonparametric models are often useful for a first analysis because they can show which items do not belong to a homogeneous set of indicators of a particular latent variable, and so can be discarded. If a parametric model does not fit the full set of items that are supposed to form a scale, the researcher typically performs a "top down" or "backward elimination" procedure to find a maximal subset of items that does fit. However, both Mokken's nonparametric cumulative model and van Schuur's nonparametric unfolding model follow the opposite strategy of a "bottom up" or "forward inclusion" procedure, called "multiple unidimensional scale analysis" by Mokken. This procedure allows the user to find two or more (possibly overlapping) subsets of items, each of which conforms to the requirements of an ordinal scale. In exploratory data analysis, it is often advisable to start with such ordinal procedures before applying the stricter parametric models, which generally fit survey data more poorly.

In this article, I am concerned with an unfolding model that assumes an ordinal latent variable. Although it is historically incorrect (cf. Coombs 1964; Coombs and Smith 1973; Leik and Matthews 1968), many textbooks suggest that unfolding is only useful for analyzing full rank orders of preference data (e.g., McIver and Carmines 1981). The best known developments of unfolding procedures are indeed those that use rank orders of preference data as incomplete rank orders of similarity data. In this approach the preference judgment of a subject for an item is interpreted as reflecting the similarity between the characteristics of that item and the subject's ideal item. True similarity data consist of pairwise similarities between all pairs of objects. But preference judgments from each subject give only the (rank order of) similarities between the subject's ideal and each of the items, not the similarities between the subjects or the similarities between the items.

Techniques for analyzing similarity data (especially multidimensional scaling [MDS], e.g., with the ALSCAL procedure, which has recently appeared in statistical packages; cf. Takane, Young, and De Leeuw 1977) are then used to analyze preference data. But, as shown in the references cited previously, improvements in IRT models make them now in some ways superior to the MDS approach: they allow analysis of the more common types of data rather than just full rank orders, they do not lead to degenerate

2. The Dutch interuniversity expertise center ProGamma distributes many of these models, e.g., BILOG, MSP, MUDFOLD, MULTILOG, PARELLA, PML, RIDA, and UNFOLD. Contact iec. ProGamma, P.O. Box 841, 9700 AV Groningen, The Netherlands.

parameter estimation as often as the MDS models do, and their fit measures are based on more sophisticated statistical models than those of the MDS approach. Moreover, they do not suffer from the low upper limit of analyzable subjects, which makes the MDS approach virtually useless for large data sets.

Many IRT procedures were initially designed for dichotomous data. In the last decade, new procedures have been developed for analyzing multicategory data that fit the cumulative model (see Sijtsma, Debets, and Molenaar 1990 for the ordinal latent variable case, and Masters 1982; Samejima 1969; Thissen and Steinberg 1984 for the interval latent variable case). In this article, I propose an extension of the ordinal unfolding model to multicategory data. In a number of respects this new model resembles the ordinal cumulative model described by Mokken and by Sijtsma, et al.

Most IRT models, including the one to be described, are unidimensional models. This might be regarded as a step back from multidimensional models such as the MDS approach. However, unidimensionality is not a liability but instead an asset of IRT models for the following reasons (see Jacoby 1991 for a discussion of this topic). First, a unidimensional model does not have the rotational indeterminacy that makes the interpretation of the different dimensions of the multidimensional Euclidean MDS model somewhat arbitrary. Second, there is no logical reason that each item ought to have a value on each dimension (note that, in MDS models, the value 0 does not mean "absence" but, rather, "centroid position"). Third, even if different dimensions are used to represent criteria for preferences for a number of items, there is no logical reason that subjects should necessarily apply these dimensions simultaneously, as the multidimensional model suggests, rather than hierarchically.

Even if preferences for a number of items are assumed to be representable by the multidimensional scaling model, this assumption can be tested through an empirical comparison with a multiple unidimensional scaling model. The MUDFOLD model, as I will explain, is a model for multiple unidimensional unfolding that uses a procedure in which maximal subsets of items are sought from a larger pool of items, with each subset forming a unidimensional unfolding scale. It shares with MDS models the assumption that researchers do not need to know in advance what unfolding dimensions are present in their data. Rather, the results from the analysis help the researcher to interpret the criteria subjects used in responding to the items.

I will describe the MUDFOLD model through a comparison of two ideal types of models: the deterministic model, in which the data contain no violations of the unfolding model, and the statistical independence model, in which the data contain as many model violations as would be expected if all responses are completely independent. The two models are linked in terms of a measure of homogeneity. I will describe the deterministic model first.

The Deterministic Unfolding Scale for Dichotomous Data

An unfolding scale consists of a set of items and a set of subjects ordered along a latent variable. The term *latent continuum* is used for the graphic representation of this variable. According to the deterministic unfolding model for dichotomous data, subjects only respond positively to items that are represented close to their position on the latent continuum, but negatively to items that are far away from that position. Since they respond positively to all items close to their own position, they will respond positively only to items that are adjacent to each other.

Let us take an example from an investigation of European party activists (Reif, Cayrol, and Niedermayer 1980; van Schuur 1984) in which the question was asked: "Are you in favor of the following proposed political measures?" Agreement with the proposed measure was indicated on a five-point scale (very much in favor, in favor, undecided, against, and very much against). The data were recoded dichotomously. In dichotomous data, one of the two response categories is called the "positive" or "scale" response, and the other response is called the "negative" response. In this data set, "very much in favor" is taken as the positive response, and the other responses were all recoded as the negative response. Eight of the proposed measures were:

A. We should reduce income differences.
B. There should be far more active control over activities of multinational corporations.
C. Greater effort should be made to protect the environment.
D. We should fight against unemployment.
E. We should fight against inflation.
F. The most severe penalties should be introduced for acts of terrorism.
G. We should reduce the capacity of public control over private enterprises.
H. Military expenditures should be increased.

If these eight items form a deterministic unfolding scale, the response patterns for the eight items combined should look like the response patterns shown in table 1. These patterns are all perfect, because the positive responses are given only to items that are adjacent to each other; i.e., we find no 0s between the 1s. Since the 1 responses show the form of a parallellogram, the unfolding analysis of dichotomous data has been called "parallellogram analysis" by Coombs (1964), the originator of unfolding analysis. This scale might be interpreted as a latent continuum measuring political ideology, going from extreme leftist to extreme rightist opinions.

An essential difference between a cumulative scale and an unfolding scale lies in the representation of the subjects along the latent continuum. In a cumulative scale, subjects are represented between the most "difficult" item to which they give the positive response and the next (still more "difficult") item to which they give the negative response. But in an unfolding scale, subjects are represented in the middle of the items they respond positively to. Note that whether certain items (e.g., survey questions) form a cumulative scale or an unfolding scale does not depend on their formulation. The same items can be used in both a cumulative scale analysis and an unfolding analysis.

Scale values are assigned to subjects and items in an unfolding scale as follows. First, the items are given a rank score in the order in which they form an unfolding scale, using only the odd numbers (how that rank order is determined will be discussed subsequently). Second, as their scale value, subjects are assigned the median value of the item values of the items to which they responded positively. An example of the assignment of scale values is shown in table 1.

Of these four subjects, subject 1 has the lowest scale value (5) and subject 4 the highest (11). Scale values of subjects can be regarded as their values on the new (latent) variable, and they can be used in further analyses just like any other dependent or independent variable.

In this article I discuss only dichotomous "pick any/*n*" data. "Pick any/*n*" data are data in which subjects may give the positive response to any of the *n* items, ranging from 0 to *n*. In this sense, responses to different items are independent. However, the procedure I will describe has also been adapted for use with "pick *k*/*n*" data (van Schuur 1984). For "pick *k*/*n*" data, each subject picks exactly the same number—*k*—of most preferred items among the total of *n* items. Full or partial rank orders of preference are an important source of such data, since they can be recoded to the *k* most preferred items and the *n-k* least preferred items. The software available for analyzing "pick any/*n*" data can also be used for analyzing "pick *k*/*n*" data. However, the generalization from analyzing "pick *k*/*n*" to "rank *k*/*n*" data is

TABLE 1. Scale Values of Subjects and Items in an Unfolding Scale

	Unfolding Scale								
Subject	A	B	C	D	E	F	G	H	Subject Value
1	1	1	1	1	1	0	0	0	5
2	0	1	1	1	1	1	0	0	7
3	0	0	1	1	1	1	1	0	9
4	0	0	0	1	1	1	1	1	11
Item value	1	3	5	7	9	11	13	15	

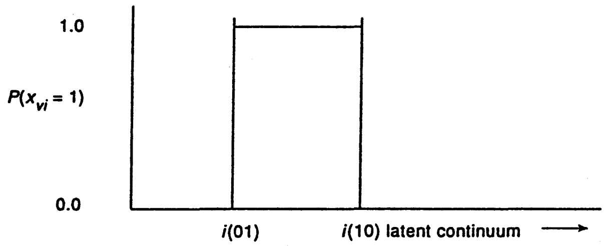

Fig. 2. Item trace line of an item that conforms to the deterministic unfolding model

not attempted in this article (but see van Blokland-Vogelesang 1991 for a nonparametric unidimensional unfolding model for full rank-order data).

The deterministic unfolding model for dichotomous "pick any/*n*" data is illustrated in figure 2. Subjects will respond positively to an item if their scale value is close to the item's scale value on the latent variable, and they will respond negatively if their scale value is either much lower or much higher. The probability function of a positive response to the item by subjects with different scale values on the latent variable is called a trace line or an item characteristic curve (ICC).

ICCs for items in the deterministic unfolding model can be regarded as a step function with two steps: the left-sided step, in which the probability of a positive response increases from 0 to 1, and the right-sided step, in which the probability decreases from 1 to 0. The two steps for each item divide the latent continuum into three areas with the responses 0, 1, and 0 (see fig. 3). For each item *i*, the three areas are separated by the two item steps, which are denoted as $i(01)$ and $i(10)$. The response value 1 means that one item step is passed, and the value 0 means that either 0 or 2 item steps are passed. For perfect unfoldable data, we can tell from the context which of the two is the case. If a positive response (1) is given to other items that are represented to the *left* of the item with a 0 response (e.g., ABC 110), then the subject should be represented to the left of both item steps of that item (C, in this case). On the other hand, if a positive response (1) is given to other items that are represented to the *right* of the item with a 0 response (e.g., ABC 011), then the subject must have passed both item steps of that item (A, in this case).

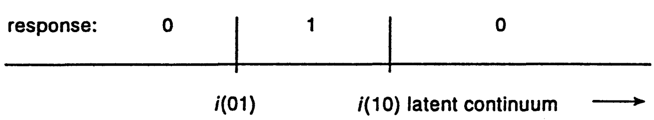

Fig. 3. Unfoldable dichotomous item with two item steps

The assumption that the items are ordered along the latent continuum implies that the order of the left-sided steps (the 01 steps) is the same as the order of the right-sided steps (the 10 steps) for all items. The response value 1 is given to item i by subjects who are represented in an area on the latent continuum between the item steps $i(01)$ and $i(10)$; the response value 0 is given by subjects outside that area. Since each item has two item steps, three items have six item steps. These divide the latent continuum into 7 (6 + 1) areas, which are rank ordered from 0 to 6. The first area, to the left of the first item step, as well as the last area, to the right of the last item step, correspond to the response pattern 000. A subject who does not respond positively to any of three items should therefore be represented in either the first or the last area. Since we have no additional information to help us choose between these two options, such a subject will be considered to have a missing datum on this latent variable.

The responses to three items, i, j, and k, in which at least one item is given the positive (1) response correspond uniquely to the five intermediate areas in which a subject can be represented, as is shown in figures 4a through 4c. In general, an unfolding scale of p dichotomous items divides the latent continuum into $2p + 1$ areas, thereby creating a new latent variable with $2p + 1$ ordered values, of which the intermediate $2p - 1$ areas are useful, but the two outer 00 . . . 0 areas are not.

The assumption of unambiguous item ordering does not imply that the distances between the items are equal, as can be seen in figures 4a, 4b, and 4c. These figures show three items, i, j, and k, which form an unfolding scale in this order. In each of the three subfigures (4a, 4b, and 4c) the order of the left-sided steps of the three items $[i(01) < j(01) < k(01)]$ is equal to the order of their right-sided steps $[i(10) < j(10) < k(10)]$. Nevertheless, even if the items have the same order for all subjects, the order of all item steps is not completely specified. This means that the rank numbers of the seven areas in the three subfigures do not necessarily correspond to the same response patterns to the three items. Although areas 0, 1, 5, and 6 always need to correspond uniquely to one response pattern (000, 100, 001, and 000, respectively), this is not true for the intermediate areas. Area 3 corresponds to either pattern 111 or 010. Areas 2 and 4 may correspond to response pattern 000, which is another indication of the uselessness of the 000 pattern.

Subjects' scale values can now be identified as the rank number of the area that is specified by their response pattern. In the deterministic model, each perfect response pattern that contains at least one positive response (1) can be assigned a unique scale value. Only the 000 pattern, the pattern with no positive response, cannot be given a unique scale value.

The scale value of subjects with a perfect response pattern is determined by counting the number of item steps they have passed. The response value 1

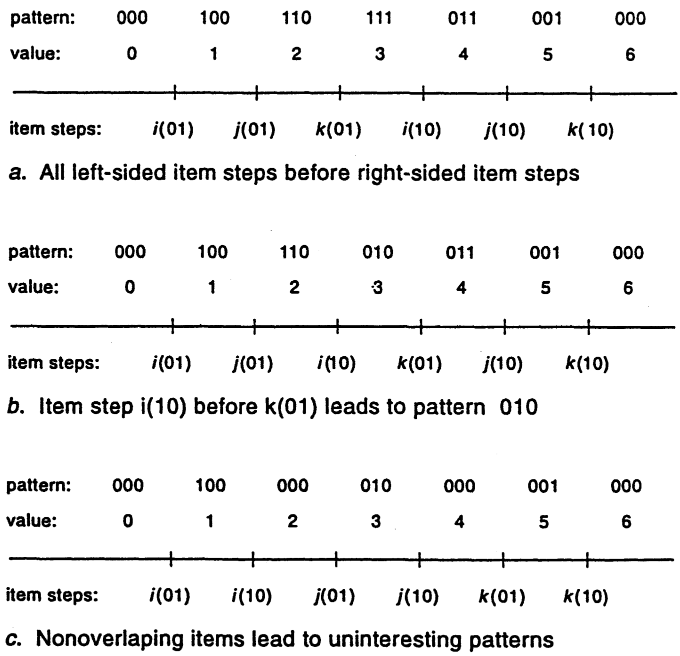

Fig. 4. Three examples of three dichotomous deterministic unfoldable items with different item step orders
***a.* All left-sided item steps before right-sided item steps**
***b.* Item step *i*(10) before *k*(01) leads to pattern 010**
***c.* Nonoverlapping items lead to uninteresting patterns**

to an item means that 1 item step has been passed (the 01 step) but not the other (the 10 step). The response value 0 means that either 0 or 2 item steps have been passed. If we know the order in which the items form an unfolding scale (see below), the response value 0 to a particular item can be interpreted: it means that 0 item steps have been passed if the responses to items to the *left* of that item contain one or more 1 responses, and it means that 2 item steps have been passed if the responses to items to the *right* of that item contain one or more 1 responses. For example, the scale value of a subject with the response pattern 110 is 2 because the total number of item steps passed is 2 (1 + 1 + 0). Similarly, the scale value of a subject with the response pattern 011 is 4, because the number of item steps passed is 4 (2 + 1 + 1). This way of calculating scale values is identical to the procedure described previously, in which the items are given the odd-valued rank scores and the subjects are assigned the median value of the rank scores of the items to which they responded positively.

Model Violations of the Deterministic Unfolding Scale

Violations of a deterministic unfolding scale are defined on the basis of a subject's responses to a triple of items that is ordered along the unfolding scale. Such a triple violates the unfolding scale if a person gives the positive response to the two outer items, but the negative response to the middle one. The response pattern 101 to the three items is thus a model violation or an error. A triple of items is the smallest set of items that can violate the unfolding model. It is therefore called the smallest scale.

This definition of error can be explained by invoking the assumption that the ICCs of a set of items are such that the order of the left-sided steps of all items is the same as the order of the right-sided steps. If a response pattern to three items, A, B, and C, is 101, the subject has responded positively to item A. If the response to item B is negative, the subject must be represented *to the left* of the left-sided item step of item B. Since item B is represented to the left of item C, the subject should also be represented to the left of the left-sided item step of item C, so the response to item C should be negative as well. But it is not. Conversely, since the response to item C is positive, and the response to item B is negative, the subject should apparently be represented *to the right* of the right-sided item step of item B. And since item B is represented to the right of item A, the subject should also be represented to the right of the right-sided item step of item A, so the response to item A should be negative. Third, since items A and C both get a positive response, the subject should apparently be represented between the 01- and the 10- item step of these two items. But this implies that the subject should also be represented between the 01- and 10- item steps of item B as well and also give the positive response to item B. It is clear, then, that a positive response to the two outer items implies a positive response to the middle item. The 101 response pattern to an ordered triple of items can thus be regarded as a violation of the model.

In an unfolding scale of more than three ordered items, a subject's response pattern contains as many errors as there are triples in which the two outer items are responded to positively and the middle one negatively. For instance, the 1001 response pattern to an ordered quadruple of items violates the model for two of the four three-item scales in this quadruple; it therefore contains two model errors. Some "wrong" response patterns thus contain more errors than others. In an unfolding scale of four items, 5 of the 16 possible response patterns contain one or more model violations: ABCD, 1001 has 2 errors (triples ABD and ACD); ABCD, 1101 has 2 errors (triples ACD and BCD); ABCD, 1011 has 2 errors (triples ABC and ABD); ABCD, 0101 has 1 error (triple BCD); and ABCD, 1010 has 1 error (triple ABC).

Note that, in the present model, if six items out of a seven-item unfolding scale receive a positive response (i.e., a score of 1), it makes a difference to the number of model violations *which* item receives the negative response (a score of 0): ABCDEFG, 1110111 contains 9 errors (ADE, ADF, ADG, BDE, BDF, BDG, CDE, CDF, CDG); ABCDEFG, 1101111 has 8 errors (ACD, ACE, ACF, ACG, BCD, BCE, BCF, BCG); and ABCDEFG, 1011111 has 5 errors (ABC, ABD, ABE, ABF, and ABG). In this the model contrasts with that of Leik and Matthews (1968), who assign to all these patterns the same number of errors, namely, one, since in each case only one change is needed to turn the pattern into a perfect response pattern.

The Statistical Independence Model and the Coefficient of Homogeneity

In previous scaling procedures, particularly Guttman's original proposal, the number of errors observed was compared to the maximum number of errors possible in the data set. The purpose of that procedure was to establish whether the scaling model fit the data sufficiently. However, if we think of establishing a criterion of scalability in terms of hypothesis testing, we generally compare the hypothesis that our set of items do indeed form an unfolding scale against the null hypothesis. But what is the null hypothesis in this situation?

When we compare the number of errors observed to the maximum number of errors, we are saying, in essence, that the null hypothesis is that the data conform to a model that is as different from an unfolding model as possible. However, Loevinger (1948) proposed that a more appropriate null hypothesis is that the items are independent: they neither form an unfolding scale nor conform to some other extreme model. In the case of independence, we cannot predict from a positive response to one item whether the subject will give the positive response to an adjacent item as well. Loevinger (1948) formulated this null hypothesis in terms of the coefficient of homogeneity, *H:*

$$H = 1 - \frac{E(\text{obs})}{E(\text{exp})},$$

where $E(\text{exp})$ indicates the number of errors expected under statistical independence.

Mokken (1970) reintroduced this coefficient as a criterion for scalability for his nonparametric cumulative scaling procedure. His suggestion is also adopted for the unfolding procedure under discussion.

The maximum possible value of the H-statistic, 1, is unambiguously interpretable: it means that there is no violation of the model. The value 0 is also unambiguous: it represents the amount of homogeneity in a data set consisting of statistically independent responses. A statistical test applicable to the H-coefficients in the unfolding procedure was developed by Post (1988) to compare the null hypothesis $H = 0$ in the population with the alternative one-sided hypothesis $H > 0$. In this test, the H-coefficient is transformed to a z-score in a standard normal distribution with mean 0 using a calculated standard deviation. If $z > 1.64$, the probability of finding a scale in a random data set (i.e., with $H = 0$) is smaller than 5 percent (one-tailed test). In the unfolding procedure, z is called the t-statistic. My experience with a large number of applications, along with simulation results, suggests that H-values over 0.30 are generally not only statistically significant but also substantively relevant for the interpretation of a set of items as a homogeneous scale. H-values over 0.30 are therefore interpreted as indicating a good fit to the model.

The degree of homogeneity can be calculated for the smallest scale as $H(ijk)$, for each individual item as $H(i)$, and for the whole data set as H. The smallest unfolding scale consists of three items—say A, B, and C—that are unfoldable in this order. The amount of observed error, or E(obs), for item triple (A,B,C) is the frequency with which items A and C get the positive response but item B gets the negative response from the same subject. The amount of expected error, or E(exp), is calculated as

$$p(A) \cdot [1 - p(B)] \cdot p(C) \cdot N,$$

in which $p(A)$, $p(B)$, and $p(C)$ are the relative frequencies with which the positive response to these items was given, and N is the sample size of the data set.

The values of E(obs) and E(exp) for the whole scale are found simply by summing the E(obs) and E(exp), respectively, of all smallest scales in the data. The homogeneity of an individual item i is found by summing E(obs) and E(exp), respectively, of only those smallest scales that contain the item. Post's t-statistic can be used for testing the null hypothesis that $H(ijk) = 0$, that $H(i) = 0$, or that $H = 0$. A small numerical example of the calculation of H, $H(i)$, and $H(ijk)$ from E(obs) and E(exp) is given in the Appendix.

Roskam and his co-workers (Roskam, Van den Wollenberg, and Jansen 1986) have objected to the use of the coefficient of homogeneity in Mokken's nonparametric cumulative scaling procedure on grounds that the H-value depends on the distribution of the population along the latent continuum. If a population is very homogeneous (i.e., its distribution has a small standard deviation), the H-coefficients will also be very low. This criticism would also apply to its use in the present unfolding procedure. However, Mokken, Lewis, and Sijtsma (1986) have replied that the dependence of the H-coefficients on the homogeneity of the population is an advantage rather than a disadvantage in Mokken's model, and their rebuttal also holds for the unfolding analysis. I agree with Mokken et al., and would only add that whatever the reason for a low H-coefficient, this outcome indicates that the scale is useless for discriminating among the sample of subjects, and it is a desirable property of a scaling procedure to discard useless scales.

Searching for the Maximal Scalable Subset of Items

A nonparametric cumulative or unfolding scale is found with a bottom-up search procedure that begins either with a researcher-defined ordered start set of items or with the best smallest scale. Researchers can select a start set (of at least three items) on the basis of any theory they might have; this option will not be discussed further here. The best smallest scale is a triple of items selected by the computer from among all possible ordered triples in accordance with the following requirements.

1. Its $H(ijk)$-coefficient of homogeneity is significantly larger than 0.
2. Its $H(ijk)$-coefficient is larger than a prespecified lower boundary. As a lower boundary, the value 0.30 is generally used.
3. Among the smallest scales that conform to these two requirements, the unfolding scale should conform to two more specific requirements.
   3.1. The best smallest unfolding scale has to conform to requirements 1 and 2 in only one of its three possible permutations (reflections are admissible transformations). In the cumulative scaling model we know the order of the items on the basis of the frequency with which the positive response to each of the items is given. In the unfolding model, in contrast, we also need to find the best—and preferably unique—order in which the items form an unfolding scale. The best smallest unfolding scale is therefore a unique triple, that is, a triple of items that has a positive $H(ijk)$ value in one permutation but a negative H-value—$H(jik)$ and $H(ikj)$—in the other two permutations.
   3.2. Among the smallest scales that conform to the three requirements above, we have two options for selecting the best triple.
      i. We select as the smallest scale the ordered triple (i,j,k) for which the sum of the frequencies of the observed patterns 111, 110, and 011 is highest. This prevents the procedure from finding a best smallest scale for which both E(obs) and E(exp) are very low, since such a scale will not be useful for the purpose of measurement.
      ii. We select as the smallest scale the ordered triple (i,j,k) that has the highest t-value (the most significant $H(ijk)$-value) in

> the scale permutation, provided that its t-values are significantly lower than 0 in both nonscale permutations.

Once the best smallest scale is found, new items are added to the scale one by one, as long as the first two requirements continue to be met (the H-coefficient of each new item is statistically significantly larger than zero, and higher than a prescribed lower boundary). Among all the items that conform to these requirements at any point in the procedure, the next item to be selected is the one that gives rise to the highest H-value of the unfolding scale as a whole.

In selecting items to add to an existing unfolding scale an additional requirement can optionally be used: from those items that conform to requirements 1 and 2, the item is chosen that gives rise to the highest H-value of the unfolding scale as a whole and that can be represented in only one position in the scale. In practice, it often happens that an additional item can be represented in more than one position in the scale. In this case, the procedure selects the item that can be represented in the smallest number of adjacent positions. The Appendix gives an example of this selection procedure for a small unfolding scale of four items.

My experience with this procedure shows that, in most unfoldable data sets, there are a number of unique triples from among which a best smallest scale can be selected. In the rare case in which no unique triples are found, researchers can use additional information about the triples (e.g., which ones are most popular, which ones appear more often than expected under statistical independence, etc.) to guide them in defining a small, ordered start set of items to begin the search procedure.

The selection procedure might be suspected of being either too strict or too lenient. It would be too strict if we insisted that each item have a unique position in the unfolding scale, since we would then find very few scales with six or more items. But it would be too lenient if we allowed a new item to be included in a scale if it is admissible in many combinations of positions. Experience suggests that MUDFOLD's intermediate strategy gives interpretable results.

In addition, the insistence that each triple of unfoldable items must have a positive $H(ijk)$-value turns out to be a remarkably stringent requirement. This is especially true for long scales, because a scale of p items contains $p(p-1)(p-2)/6$ triples. It is also the case for triples in which the number of errors is very small (e.g., with a low frequency of observed errors—say 1, 2, or 3—and a frequency of expected errors of, say, 0.9, 1.8, or 2.7), in which the t-statistic still suggests a significant deviation from $H(ijk) = 0$. In such cases, a pragmatic solution is to allow a few negative $H(ijk)$-values, as long as the $H(i)$-value of all items is still above the relevancy lower boundary of 0.30.

Scale Values of Subjects with Imperfect Response Patterns

Scale values for subjects and items were defined in a previous section for the deterministic, or perfect, unfolding scale. But how are the scale values of subjects defined in the presence of model violations?

For imperfect response patterns—those in which the 0 (negative) response occurs between two or more 1 responses—a decision must be made about whether the 0 response means 0 or 2 item steps have been passed. This decision is based on the relative number of 1 responses to the left and right of the 0 response in an ordered response pattern. If the *majority* of 1 responses is to the left of the 0 response, we assume that item step (01) has not been passed. Conversely, if the *majority* of 1 responses is to the right of the 0 response, we assume that both item steps (01) and (10) have been passed. Finally, if the number of 1 responses to the left and the right of the 0 response is the same, we assume that one item step has been passed. An example is given in table 2.

The procedure for establishing scale values for imperfect patterns is identical to the procedure described previously, which takes as the scale value the median of the rank scores—specified only with odd numbers of the items to which a positive response is given. This scale value is assigned to respondents regardless of the number of errors in their response patterns. Subjects who do not respond positively to any of the items are not assigned a scale value. Instead, their responses are defined as missing data.

TABLE 2. Scale Values for Perfect and Imperfect Response Patterns in Terms of the Number of Item Steps Passed and the Number of Errors in Terms of the Number of 101 Triples in the Response Pattern

	Item					Number of Item Steps Passed						
Subject	A	B	C	D	E	A	B	C	D	E	Scale Value	Number of Errors
1	1	1	0	0	0	1	1	0	0	0	2	0
2	0	1	1	1	0	2	1	1	1	0	5	0
3	0	0	0	0	1	2	2	2	2	1	9	0
4	1	0	1	1	1	1	2	1	1	1	6	3
5	1	1	1	0	1	1	1	1	0	1	4	3
6	1	1	0	1	1	1	1	1	1	1	5	4
7	1	0	0	1	1	1	2	2	1	1	7	4
8	1	0	1	0	1	1	2	1	0	1	5	4

Some Additional Goodness-of-Fit Diagnostics

The H-coefficients for triples of items, single items, and the whole scale are important measures of fit. But there are additional diagnostics that can help us determine whether the unfolding model is appropriate to the data. These diagnostics do not have a firm statistical foundation, but they can be used to "eyeball" deviations from the model. Work is in progress to develop statistical criteria for these diagnostics.

A first diagnostic is to compare the probability of a positive response to an item for subjects from groups with different scale values. If an item indeed has a single-peaked item characteristic curve, then the order of probabilities of a positive response to the items in their unfoldable order should also be single-peaked, both rowwise and columnwise. The highest probability should shift from the first to the last row, as we move from left to right across the columns, and it should shift from the first to the last column, as we move from top to bottom along the rows. This can simply be checked by the table that gives the probability of positive response to each item for subjects with a given scale value.

As an example to illustrate this and the following diagnostics, I take the attitudes of 330 Republican and Democratic party activists toward five U.S. politicians (see table 3). The data, taken from Rapoport, Abramowitz, and McGlennon 1986, were sympathy scores given at the Missouri state convention in 1980 for five politicians. Subjects with missing data, as well as subjects who gave a favorable response to none, only one, or all of the five politicians were excluded. The original values have been recoded as a dichotomy: 1 = (very) favorable; 0 = neutral or unfavorable. The five politicians were: A = Carter, B = Kennedy, C = Anderson, D = Reagan, and E = Bush. Their unfoldability will be demonstrated more extensively in the next section. Here it is enough to note that positive attitudes for Kennedy, Carter,

TABLE 3. Percentages of Positive Scores to Each Unfoldable Item by Subjects with Different Scale Values

Scale Value	N	Kennedy 0	Kennedy 1	Carter 0	Carter 1	Anderson 0	Anderson 1	Bush 0	Bush 1	Reagan 0	Reagan 1
1–2	31	10	90	0	100	100	0	100	0	100	0
3	36	36	64	22	78	31	69	94	6	97	3
4	30	67	33	30	70	30	70	67	33	100	0
5–6	20	65	35	40	60	35	65	30	70	60	40
7	23	100	0	78	22	26	74	13	87	4	96
8–9	190	100	0	100	0	100	0	1	99	0	100
N	330	244	86	251	79	254	76	95	235	109	221

Anderson, Bush, and Reagan can be represented as an unfolding scale in this order.

Looking from left to right for each scale value group, the expected single peakedness of the probabilities of the 1 responses can be observed. Subjects in the lowest scale value group have the highest probability of responding positively to Kennedy and Carter. Subjects with the scale value 4 have a low probability (33 percent) of giving a positive response to Kennedy, a high probability of a positive response to both Carter and Anderson, and a low probability of a positive response to Bush and Reagan. Subjects with the scale values 5 or 6 give the highest response probability to Bush, and lower ones to politicians to the left or to the right of him. Subjects with high scale values (7 and up) give the highest response probability to Reagan, and decreasing probabilities to politicians farther away from him.

Looking from top to bottom in the columns for the probabilities of the 1 responses to each of the items, the expected single-peaked pattern is also found, with only a few deviations (e.g., 65 percent of the subjects with scale value group 5–6 are favorable toward Anderson, but at least 70 percent would have been expected from the preceding and following rows).

A second diagnostic is found in the characteristic monotonicity pattern that is expected in the dominance matrix of a set of unfoldable items. The dominance matrix is the square matrix in which rows and columns are made up of the items in their unfoldable order, and in which cell i,j contains the percentage of subjects who gave the positive response to the row item i and the negative response to the column item j. Under the deterministic unfolding model, this percentage should be low for adjacent items, but should increase with the distance between items. The dominance matrix should thus show for each row that the percentages decrease from the leftmost column to the diagonal, and increase from the diagonal to the rightmost column. Such a monotonicity pattern is called "characteristic monotonicity." Table 4 shows an example in which cell (i,j) contains the percentage of subjects who give the positive response to item i but the negative response to item j.

TABLE 4. Dominance Matrix for Five Unfolding Items

	B	A	C	E	D
(B) KENNEDY	—	11	13	22	23
(A) CARTER	13	—	15	18	23
(C) ANDERSON	12	12	—	15	17
(E) BUSH	69	63	63	—	8
(D) REAGAN	66	64	61	3	—

Note: This dominance matrix conforms to the expected pattern of characteristic monotonicity.

TABLE 5. Adjacency Matrix for Five Unfolding Items

	B	A	C	E	D
(B) KENNEDY	—				
(A) CARTER	13	—			
(C) ANDERSON	10	10	—		
(E) BUSH	2	7	8	—	
(D) REAGAN	1	3	6	63	—

Note: This adjacency matrix conforms to the expected pattern of decreasing proportions from the diagonal to the left and down.

A third diagnostic, similar to the second, is found in the monotonicity pattern of the adjacency matrix. The adjacency matrix is the square matrix in which rows and columns are made up of the items in their unfoldable order and in which cell i,j contains the percentage of subjects who give the positive response to both items i and j. Under the deterministic unfolding model, this percentage should be high for adjacent items but should decrease with increasing distance between the items. The adjacency matrix in table 5 should thus show that the percentages decrease from the diagonal both to the left on the same row and down on the same column.

Post (1992) has formulated an explicit nonparametric probabilistic version of the unidimensional unfolding model for dichotomous data. She shows that, in her model, the expectation that the adjacency matrix should show a monotonicity pattern does not necessarily hold for some subject distributions. As an alternative diagnostic, she introduces the conditional adjacency matrix. Like the dominance and adjacency matrices, the conditional adjacency matrix is a square matrix in which rows and columns contain the items in their unfoldable order. In a conditional adjacency matrix, the value of each cell (ij)—the number of subjects who give the positive response to both items i and j—is specified relative to the number of subjects who give the positive response to the row item (item i). Thus, each cell contains the value $f(ij)/f(i)$. Post showed that a nonparametric probabilistic unfolding scale with isomorphic trace lines gives rise to a conditional adjacency matrix with the property that the maximum cell value is found in cells that "move" from top left to bottom right in the matrix. The combination of maximum values in cells (i,j) and cell $(i + x, j - y)$ $(x,y > 0)$ indicates a model violation. Table 6 shows the results of the conditional adjacency matrix for the attitude scores toward the five politicians. For quick eyeballing, the conditional adjacency matrix is summarized with an asterisk (*) in each row for the cell that contains the highest value. If this value is adjacent to the diagonal, then the diagonal also contains an asterisk. In a good unfolding scale, the asterisks should form

TABLE 6. Conditional Adjacency Matrix for Five Unfolding Items

	B	A	C	D	E
(B) KENNEDY	—	0.51	0.47	0.03	0.02
	—	(.054)	(.057)	(.011)	(.009)
(A) CARTER	0.56	—	0.46	0.11	0.05
	(.056)	—	(.057)	(.020)	(.015)
(C) ANDERSON	0.41	0.46	—	0.11	0.09
	(.053)	(.056)	—	(.021)	(.019)
(E) BUSH	0.09	0.30	0.36	—	0.95
	(.032)	(.050)	(.055)	—	(.015)
(D) REAGAN	0.05	0.13	0.26	0.89	—
	(.026)	(.036)	(.051)	(.020)	—

Notes: Each cell contains the proportion $f(ij)/f(i)$. The standard error of that proportion is shown in parentheses.

a pattern from top left to bottom right. Table 7 shows that this is indeed the case for the responses to the five politicians.

Post also showed that, for a nonparametric probabilistic unfolding model with isomorphic trace lines, the correlation matrix of the items, represented in their order on the unfolding scale, should exhibit two sign changes at most: from negative to positive to negative. This is a weaker form of Davison's (1977) requirements for a metric unidimensional unfolding scale, according to which the correlations should form a simplex pattern running from highly positive to zero to highly negative. Table 8 shows that the correlation matrix of the responses to the five unfolding items indeed conforms to this expected pattern.

The present software for the MUDFOLD model for dichotomous data will calculate the dominance matrix, the unconditional and conditional adjacency matrices, and the correlation matrix as additional diagnostics.

TABLE 7. Summary of Conditional Adjacency Matrix for Five Unfolding Items

	B	A	C	E	D
B	*	*			
A	*	*			
C		*	*		
E				*	*
D				*	*

TABLE 8. Correlation Matrix for Five Unfolding Items

	B	A	C	E	D
(B) KENNEDY	1.00	0.38	0.30	−0.77	−0.74
(A) CARTER	0.38	1.00	0.25	−0.54	−0.68
(C) ANDERSON	0.30	0.25	1.00	−0.43	−0.47
(E) BUSH	−0.77	−0.54	−0.43	1.00	0.75
(D) REAGAN	−0.74	−0.68	−0.47	0.75	1.00

Note: Cell (i, j) contains the product-moment correlation (phi-coefficient) of the dichotomous variables i and j mentioning/not mentioning item i and item j.

Extension of Unfolding to Multicategory Data

The unfolding procedure described here can easily be extended from dichotomous to multicategory data. Subjects who give a positive response to a dichotomous item can be represented on the latent continuum in the area between the two item steps (01 and 10). Similarly, subjects who give the highest response value (say, r) to a multicategory item (i.e., an item with response values between 0 and r) can unequivocally be represented between the two item steps ($r - 1$, r and r, $r - 1$). Subjects who give any other response can be represented either to the left or to the right of this area. As with a 0 response to a dichotomous variable, the surrounding context—that is, the items that are given the highest response—determines how a lower response should be interpreted.

To illustrate, let us take the question "How favorable are you toward Politician P?" with the response categories very favorable (code 2), somewhat favorable (code 1), and not favorable (code 0). When an item has three sponse categories (0, 1, and 2), of which 2 is the most positive, the responses can be represented along a latent continuum as shown in figure 5.

As in the case of dichotomous items, we assume that there is an unambiguous order in which the items form an unfolding scale. An unambiguous order of the items implies that the order of the item steps for each pair of adjacent response categories is the same for all items. So, if three items i, j, and k form an unfolding scale in this order, then this implies that $i(01) < j(01) < k(01)$, as well as $i(10) < j(10) < k(10)$, but also $i(12) < j(12) < k(12)$, and $i(21) < j(21) < k(21)$. Figure 6 shows some possible response patterns in a perfect deterministic unfolding scale.

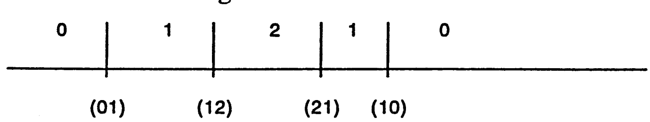

Fig. 5. Unfoldable three-category item with four item steps

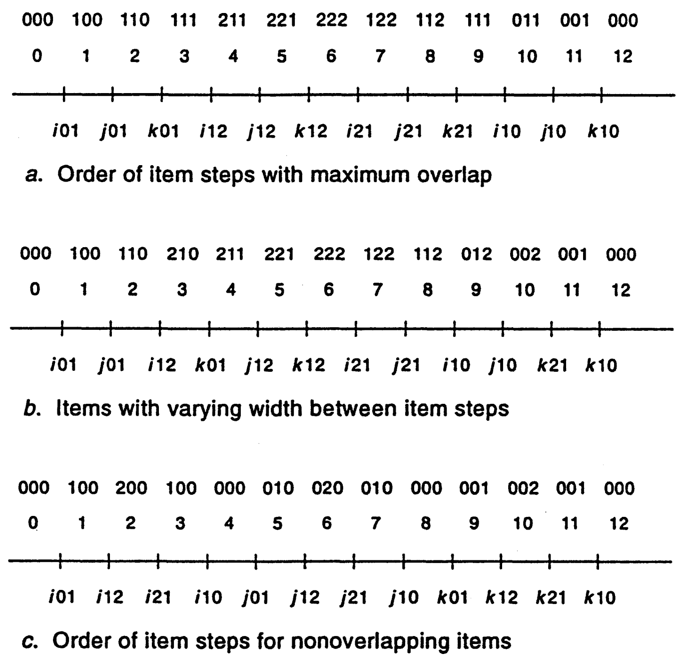

Fig. 6. Three examples of three deterministic unfoldable items with three categories each that differ in their item step orders
***a.* Order of item steps with maximum overlap**
***b.* Items with varying width between item steps**
***c.* Order of item steps for nonoverlapping items**

All three parts of figure 6 show unfolding representations that conform to the requirement that the order of the items along the latent continuum is unambiguous: for each step between adjacent item values, the order is the same: a particular step comes first for item *i*, then for item *j*, and last for item *k*.

Each item with response categories between 0 and r has $2r$ different item steps. Therefore, an unfolding scale with three items, each with three categories (0,1,2), has $3 \cdot 2 \cdot 2 = 12$ different item steps, and so divides the latent continuum into $12 + 1 = 13$ different areas. How can we assign scale values to each of these areas?

The answer is analogous to the answer in the case of dichotomous data. The highest response to an item r, (e.g., 2) is unambiguously located between the r^{th} and the $r + 1^{\text{th}}$ item step. Thus, a subject who gives the highest response r has passed r item steps. A subject who gives a lower response, t ($t < r$), has passed either t or $2r - t$ item steps, depending on the context (i.e., which items have been given the highest response value). As can be seen from figures 6a through 6c, it is possible to determine unambiguous scale values only for response patterns that contain the highest response category at least once.

The third example, figure 6c, shows that response patterns that do not contain the highest response category can appear at different locations along the latent continuum. It is therefore prudent to discard subjects who do not give the highest response value to at least one of the unfoldable items; this is done by substituting a missing value for their scale value. For subjects who give the lowest possible response value to all items, such a procedure is obvious: they would otherwise be assigned either the lowest or the highest scale value. Since we have no information which of these two values it should be, we might make serious errors in the assignment of such scale values.

Model violations for multicategory data are similar to those in dichotomous data. Suppose that three items i, j, and k form a deterministic unfolding scale in this order, and that a subject gives the response value t to item i, u to item j, and v to item k (e.g., *ijk*, 210). Each response pattern in which u is smaller than either t or v is a model violation. The logic is as follows: if $t > u$, then the subject has not yet passed the item step $(t - 1, t)$ for item j, and so must also not have passed it for item k. The response to item k must therefore be, at most, not higher than the response value u. A value $v > u$ therefore indicates a model violation. Conversely, $u < v$ implies that the subject has already passed the item step $(u, u - 1)$ for item j, and therefore must also have passed it for item i. The response to item i must therefore be, at most, no higher than the response value u. A value $t > u$ therefore indicates an analogous model violation.

Do all erroneous response patterns violate the unfolding model to the same extent? The answer is no. The response pattern *ijk*, 202, for instance, is more in error than the response pattern *ijk*, 212. This can be shown by considering the implications for the response value of the third item when we accept the response values for the other two. As an example, I use the observed response pattern *ijk*, 202. This response pattern will be called the explicit error. The response to each of the three item pairs implies additional implicit errors.

—The response pattern *ij*, 20 suggests that the subject has not yet passed item step $j(01)$ and, therefore, should also not have passed item step (01) for the third item, k. Thus, the response pattern *ij*, 20 suggests response 0 for item k, or *ijk*, 200 for the triple. Responses with a higher value for k—*ijk*, 201 or *ijk*, 202—thus violate the model.

—Similarly, the response pattern *jk*, 02 suggests that the subject has already passed item step $j(10)$ and so should also have passed item step $i(10)$. The only admissible response to item i should therefore be 0, or *ijk*, 002 for the triple. Responses with a higher value for *i*—*ijk*, 102 or *ijk*, 202—thus violate the model.

—The response pattern for the third pair is *ik*, 22. This response implies that the middle item, *j*, should also have the highest response value, 2, or *ijk*, 222 for the triple. Responses with a lower value for *j*—*ijk*, 212 or *ijk*, 202—thus violate the model.

Response pattern 202 therefore contains four errors: not only the explicit error, 202, but also three implicit errors: patterns 201, 102, and 212. Following the same logic, response pattern 212 contains only the one explicit error. In general, if t, u, and v are response categories, $u < t, v$, and tv^* is defined as the smallest value of t and v, then the response pattern *ijk*, *tuv* for the ordered triple of items, i, j, and k contains $(v - u) + (t - u) + (tv^* - u) - 2 = v + t + tv^* - 3u - 2$ errors: $(v - u)$ errors are implied by the response *ij*, *tu*; $(t - u)$ errors are implied by the response *jk*, *uv*; and $(tv^* - u)$ errors are implied by the response *ik*, *tv*. The explicit error appears in each of these three pairs and, therefore, needs to be subtracted twice.

We can calculate the total number of errors in each response pattern by summing the number of errors in each triple of items over all triples. The number of errors in the scale as a whole is simply the sum of the number of errors in each response pattern. Alternatively, we can calculate the number of errors in each triple of items and then sum across all triples. The number of errors associated with a particular item can also be found by summing the number of errors in the response patterns of the triples containing that item.

The calculation of the expected number of errors in each triple of items is similar for multicategory data and for dichotomous data.[3] For the response pattern with the values $A = t$, $B = u$, and $C = v$,

$$E(\exp) = \sum_{t,v>u} N \cdot p(A_t) \cdot [1 - p(B_u)] \cdot p(C_v).$$

The calculation of the expected number of errors in each triple of items is similar for multicategory data and for dichotomous data.[3] For the response pattern with the values $A = t$, $B = u$, and $C = v$,

The ordering of items in an unfolding scale for multicategory data is also established in the same way as for dichotomous data. For each triple of items,

3. In principle, the number of response categories should be small enough that the relative frequency of occurrence of each value of a variable can be estimated with reasonable accuracy. This requirement may sometimes pose difficulties. For example, in the case of 100-point feeling thermometer scores for politicians, the expected frequencies of each response pattern can be estimated only roughly.

the number of errors observed and the number of errors expected are calculated, and the hierarchical, bottom-up procedure described for dichotomous data is applied. This procedure finds maximal subsets of unfoldable items.

Goodness of fit for the multicategory model can be assessed with the following by now familiar techniques.

1. Once the observed and expected number of errors for each triple is established, Loevinger's H-coefficients can be determined for each triple, for each item, and for the whole scale. The requirements are that each triple in the scale should have a positive $H(ijk)$-value, and that each item should have an $H(i)$-value larger than 0.30. The whole scale then also has an H-value larger than 0.30.
2. The procedure gives the table that shows the probability of a specific response to each item for subjects with a given scale value. Probabilities for the highest (or lowest) response should follow the same characteristic monotonicity patterns as the probabilities for the 1 (or 0) response in the dichotomous case.
3. The dominance matrix is interpreted in the same way as for dichotomous data. Cell (i,j) contains the proportion of subjects who give a higher value to the row item than to the column item, so the cell values should be lowest along the diagonal and increase to both the leftmost and rightmost column.
4. The adjacency matrix and the conditional adjacency matrix can be interpreted just as for dichotomous data if the researcher specifies a cutting point that distinguishes high and low response scores. Adjacency is then defined as the proportion of subjects who give the high response score to a pair of items. The properties of the adjacency and conditional adjacency matrix are those described earlier: in the adjacency matrix cell proportions decrease for items with increasing distance, and in the conditional adjacency matrix maximal cell values "move" from top left to bottom right.
5. The correlation matrix should conform to the same requirements described for the dichotomous case: adjacent items should correlate positively, but items that are far apart in either direction may correlate negatively.

An Example of Multicategory Unfolding Analysis of Five Items

For this example, we take the same attitudes of Republican and Democratic party activists toward five American politicians. The original values have now been recoded on a four-point scale: 3 = (very) favorable; 2 = neutral; 1 =

TABLE 9. Descriptive Statistics for Responses to Five Attitude Questions

Candidate	(Very) Favorable 3	Neutral 2	Somewhat Unfavorable 1	Very Unfavorable 0
(A) Carter	.29	.04	.15	.52
(B) Kennedy	.14	.05	.11	.69
(C) Anderson	.13	.16	.22	.49
(D) Reagan	.61	.04	.04	.31
(E) Bush	.39	.19	.17	.25

somewhat unfavorable; 0 = very unfavorable. The data of 571 subjects have been used.

Table 9 shows that Reagan was favored most by the respondents and that only a small minority gave neutral responses. Kennedy was disliked the most, followed by Carter. There were 8 subjects with no favorable scores for any of the politicians; their data have been excluded from further analysis. Two hundred seventy-eight respondents gave the highest (favorable) score to one politician, 241 respondents to two, 41 to three, and, finally, 3 respondents to four politicians.

Table 10 shows the H-coefficient for each triple of items in each of its three permutations. This table contains the central information that is used in the exploratory procedure that searches for a scale.

Table 10 reads as follows. Each of the ten triples of items (from ABC to CDE) in each of its three different permutations is a candidate best smallest unfolding scale. Triple ABC, for instance, might form a scale in the orders

TABLE 10. Observed and Expected Errors and $H(ijk)$-Coefficients for All Triples of Items in Each of Their Three Permutations

	First as Middle			Second as Middle			Third as Middle		
	E(o)	E(e)	$H(jik)$	E(o)	E(e)	$H(ijk)$	E(o)	E(e)	$H(ikj)$
ABC	46	52	.12	92	102	.12	55	53	−0.03
ABD	22	76	.71	91	150	.39	103	29	−2.58
ABE	28	79	.65	113	158	.28	76	32	−1.36
ACD	126	122	−.03	51	136	.63	93	47	−1.00
ACE	137	129	−.06	58	135	.57	52	49	−0.06
ADE	305	190	−.61	79	71	−.12	14	88	0.84
BCD	153	154	.01	15	84	.82	92	30	−2.09
BCE	176	163	−.08	14	84	.83	60	30	−0.98
BDE	323	235	−.37	58	45	−.29	11	52	0.79
CDE	275	213	−.29	74	73	−.02	7	78	0.91

← Liberal Conservative →

	Kennedy	Carter	Anderson	Bush	Reagan
$H(i)$:	.68	.61	.66	.74	.77

Fig. 7. Final result of scale search procedure, $H = 0.69$

BAC, ABC, or ACB. The scales ABC and BAC each have an H-value of only 0.12, well below the recommended lower boundary of 0.30. The order ACB is even worse: its H-value is negative (-0.03). Six of the ten triples that can be constructed from items A through E are unique; that is, their H-coefficient is positive in only one of their three permutations: ACD, ACE, AED, BCE, BED, and CED. For triple ACD, for instance, only 51 subjects gave an error response: a higher response to A (Carter) and D (Reagan) than to C (Anderson). On the basis of the overall popularity of the three politicians, we would have expected 136 subjects to give an error response to this ordered triple: the H-coefficient for triple ACD is therefore $1 - 51/136 = 0.63$.

Triple AED is selected as the best unique triple because the number of subjects who gave the highest scores to the adjacent items (A and E [79] + E and D [305]) is higher than for any of the other unique triples. Next, item B is added in initial position to form scale BAED. The alternative would have been to add item C in the second position, giving scale ACED. But since the H-coefficient of scale BAED is higher than that of scale ACED (0.75 vs. 0.70), BAED is selected. Finally, item C is added in the third position. Figure 7 gives the final result of the search procedure: all five items take part in the unfolding scale BACED, which has an acceptable H-coefficient of 0.69.

Once the unfoldable order of the items is found, we can further investigate the scale's goodness of fit. Among the remaining 563 ($571 - 8$) respondents, 435 (77 percent) gave a response pattern that was perfectly compatible with the underlying unfolding scale. Twenty-one percent of the respondents had a model violation in 1, 2, or 3 of the ten triples.

The dominance matrix shown in table 11 displays a few deviations from the expected pattern of characteristic monotonicity, especially around cells AC and AE: this deviation stays well within the margins of sampling error, however.

Table 12 is a cross-table that shows, for each of the nine successive scale value groups, the percentage of respondents who give each of the response values for the five items in their unfoldable order. The patterns in this cross-table conform to the expected pattern of characteristic monotonicity. Both rowwise and columnwise the responses show a single-peaked function, with the peak moving from left to right. There are only a few deviations from the expected monotonicity patterns, and they are rather small.

TABLE 11. Dominance Matrix: Percentage of Respondents Who Gave a Higher Score to the Row Variable than to the Column Variable

	B	A	C	E	D
(B) Kennedy	—	11	14	20	23
(A) Carter	36	—	30	28	32
(C) Anderson	35	29	—	15	21
(E) Bush	63	57	53	—	16
(D) Reagan	65	62	61	28	—

Note: Cell (ij) + cell (ji) do not total 100 percent because of the number of subjects who gave both items the same score.

Discussion and Summary

The MUDFOLD procedure described in this article fills an empty cell in the classification of the scaling models that are generally known as Item Response Theory (IRT) models. The nonparametric unfolding procedure for dichotomous data has been described in previous publications (van Schuur 1984, 1987, 1988, 1989). This article improves on those publications in three ways. First, the model assumptions are better described by the introduction of item steps. Second, scale values for subjects are derived from the model in terms of these item steps. Third, the model is extended to items with more than two response categories, such as Likert scales. In many respects, the MUDFOLD procedure can be regarded as the unfolding analogue of the cumulative model developed by Mokken (1970) and extended by Molenaar (cf. Sijtsma, Debets, and Molenaar 1990).

Advantages of the unfolding procedure for multicategory data are that measurement instruments based on the unfolding model can be made more precise with fewer multicategory variables and that no information is lost by recoding the original responses. Disadvantages are that response patterns of subjects that do not contain the highest (most positive) score cannot be interpreted unambiguously. The researcher may need to recode the highest response categories to circumvent this problem.

The search procedure that finds a maximal subset of unfoldable items has the advantage of serving as a first analysis that can be used to identify those items that are probably appropriate for a parametric unfolding analysis. As with any hierarchical clustering procedure, this procedure cannot guarantee that the maximum subset found is indeed the overall best subset. The MUDFOLD procedure shares the shortcomings of most IRT models in that no provisions are yet made for missing data, for guessing, or for other problems typical of survey

TABLE 12. Cross-table of the Percentage of Respondents Who Gave a Response Value to Each of the Five Politicians by (Recoded) Scale Value Group

			Kennedy				Carter				Anderson				Bush				Reagan			
Scale Value	Recoded Scale Value	*N*	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
3–8	1	51	0	18	18	65	27	10	6	57	61	20	20	0	82	14	4	0	92	6	2	2
9	2	53	59	11	2	28	17	4	0	79	62	9	8	21	87	9	4	0	96	0	4	0
10–12	3	69	30	19	13	38	6	7	3	84	13	23	23	41	20	49	26	4	75	19	6	0
13–19	4	53	40	30	17	13	6	11	15	68	9	13	34	44	2	8	42	49	36	13	23	28
20–21	5	31	77	23	0	0	45	32	13	10	7	10	42	42	0	0	13	87	0	0	7	93
22	6	45	91	9	0	0	47	47	7	0	4	47	49	0	0	0	16	84	0	0	2	98
23	7	61	97	3	0	0	72	28	0	0	28	64	8	0	0	0	13	87	0	0	2	98
24	8	102	97	3	0	0	93	7	0	0	81	15	4	0	0	7	18	75	0	0	0	100
25–27	9	98	98	2	0	0	94	6	0	0	93	7	0	0	35	40	26	0	0	0	0	100

Note: Values are: 0 = very unfavorable; 1 = somewhat unfavorable; 2 = neutral; 3 = very favorable.

responses. Work is in progress to give a better statistical foundation for the diagnostics that are presently available.

Software for the original dichotomous model is available (see n. 1), and the extensions presented here will be incorporated shortly. It should be emphasized that the question format in which data are collected does not prescribe a specific model for data analysis. For example, many researchers believe that only rank orders of preferences should be unfolded; they assume that Likert-scale data should be factor-analyzed and dichotomous data should be subjected to Guttman scaling or reliability analysis. But this is incorrect: Likert-scale data and dichotomous data can also be unfolded. The danger of factor analyzing data that should be unfolded has been described in the mathematical psychology literature (e.g., Coombs and Kao 1960; Davison 1977; Ross and Cliff 1964), but this warning has been ignored by more substantively oriented researchers. One reason for this is probably that IRT-based software for unfolding has not been available in the past. The development of unfolding models that can analyze the types of data that are most frequently collected can therefore be regarded as an important extension of our methodological tool kit.

REFERENCES

Andrich, D. 1988. "The Application of an Unfolding Model of the PIRT Type for the Measurement of Attitude." *Applied Psychological Measurement* 12:33–51.

Böckenholt, V., and I. Böckenholt. 1990. "Modeling Individual Differences in Unfolding Preference Data." *Applied Psychological Measurement* 14:257–69.

Brady, H. E. 1990. "Traits versus Issues: Factor versus Ideal-Point Analysis of Candidate Thermometer Ratings." *Political Analysis* 2:97–129.

Coombs, C. H. 1964. *A Theory of Data*. New York: Wiley.

Coombs, C. H., and R. C. Kao. 1960. "On a Connection between Factor Analysis and Multidimensional Unfolding." *Psychometrika* 25:219–31.

Coombs, C. H., and J. E. K. Smith. 1973. "On the Detection of Structure in Attitudes and Developmental Processes." *Psychological Review* 5:337–51.

Croon, M. 1990. "Latent Class Analysis with Ordered Latent Classes." *British Journal of Mathematical and Statistical Psychology* 43:171–92.

Davison, M. 1977. "On a Metric, Unidimensional Unfolding Model for Attitudinal and Developmental Data." *Psychometrika* 42:523–48.

DeSarbo, W. A., and D. L. Hoffman. 1986. "The Simple and Weighted Thresholds Model for the Spatial Representation of Binary Choice Data." *Applied Psychological Measurement* 10:247–64.

Fischer, G. 1974. *Einführung in die Theorie Psychologischer Tests*. Vienna: Huber Verlag.

Formann, A. K. 1988. "Latent Class Analysis Models for Nonmonotone Dichotomous Items." *Psychometrika* 53:45–62.

Guttman, L. 1950. "The Basis for Scalogram Analysis." In *Measurement and Prediction,* ed. S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, and J. A. Clausen, 216–57. Princeton: Princeton University Press.

Hoijtink, H. 1990. "A Latent Trait Model for Dichotomous Choice Data." *Psychometrika* 55:641–56.

Jacoby, W. G. 1991. *Data Theory and Dimensional Analysis.* Sage University Papers Series on Quantitative Applications in the Social Sciences, no. 07-078. Newbury Park, Calif.: Sage.

Leik, R. K., and M. Matthews. 1968. "A Scale for Developmental Processes." *American Sociological Review* 54:62–75.

Loevinger, J. 1948. "The Technique of Homogeneous Tests Compared with Some Aspects of 'Scale Analysis' and Factor Analysis." *Psychological Bulletin* 45: 507–30.

McIver, J. P., and E. G. Carmines. 1981. *Unidimensional Scaling.* Sage University Papers Series on Quantitative Applications in the Social Sciences, no. 07-024. Beverly Hills and London: Sage.

Masters, G. A. 1982. "Rasch Model for Partial Credit Scoring." *Psychometrika* 47:149–74.

Mokken, R. J. 1970. *A Theory and Procedure of Scale Analysis with Applications in Political Research.* New York and Berlin: De Gruyter (Mouton).

Mokken, R. J., and C. Lewis. 1982. "A Nonparametric Approach to the Analysis of Dichotomous Item Responses." *Applied Psychological Measurement* 6:417–30.

Mokken, R. J., C. Lewis, and K. Sijtsma. 1986. "Rejoinder to 'The Mokken Scale: A Critical Discussion.'" *Applied Psychological Measurement* 10:279–85.

Molenaar, I. W. 1982. "Mokken Scaling Revisited." *Kwantitatieve Methoden* 3: 145–64.

Poole, K. T. 1984. "Least Squares Metric Unidimensional Unfolding." *Psychometrika* 49:311–23.

Post, W. J. 1988. "Distribution of the Observed Number of Errors as Defined in the MUDFOLD Model under Statistical Independence in the Pick/Any Case." In *The Many Faces of Multivariate Analysis: Proceedings of the 1988 SMABS Conference,* ed. M. G. H. Jansen and W. H. van Schuur, 2:255–69. Groningen: RION.

Post, W. J. 1992. *Nonparametric Unfolding Models: A Latent Structure Approach.* Leiden: DSWO Press.

Rapoport, R. B., A. I. Abramowitz, and J. McGlennon. 1986. *Life of the Parties.* Lexington: University Press of Kentucky.

Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Nielsen and Lydiche.

Reif, K., R. Cayrol, and O. Niedermayer. 1980. "National Political Parties' Middle Level Elites and European Integration." *European Journal of Political Research* 8:91–112.

Roskam, E. E., A. L. van den Wollenberg, and P. G. W. Jansen. 1986. "The Mokken Scale: A Critical Discussion." *Applied Psychological Measurement* 10:165–77.

Ross, J., and N. Cliff. 1964. "A Generalization of the Interpoint Distance Model." *Psychometrika* 29:167–76.

Rost, J. 1988. "Rating Scale Analysis with Latent Class Models." *Psychometrika* 53:327–48.

Samejima, F. 1969. "Estimation of Latent Ability Using a Response Pattern of Graded Scores." *Psychometrika Monograph Supplement* 4:1–100.

Sijtsma, K., P. Debets, and I. W. Molenaar. 1990. "Mokken Scale Analysis for Polychotomous Items: Theory, a Computer Program, and an Empirical Application." *Quality and Quantity* 24:173–88.

Takane, Y., F. W. Young, and J. De Leeuw. 1977. "Nonmetric Individual Differences Multidimensional Scaling: An Alternating Least Squares Method with Optimal Scaling Features." *Psychometrika* 42:7–67.

Thissen, D., and L. Steinberg. 1984. "A Model for Multiple Choice Items." *Psychometrika* 49:501–19.

Van Blokland-Vogelesang, R. 1991. *Unfolding and Group Consensus Ranking for Individual Preferences*. Leiden: DSWO Press.

Van Schuur, H. 1984. *Structure in Political Beliefs: A New Model for Stochastic Unfolding with Application to European Party Activists*. Amsterdam: CT Press.

Van Schuur, W. H. 1987. "Constraint in European Party Activists' Sympathy Scores for Interest Groups: The Left-Right Dimension as Dominant Structuring Principle." *European Journal of Political Research* 15:347–62.

Van Schuur, W. H. 1988. "Stochastic Unfolding." In *Sociometric Research*, vol. 1, *Data Collection and Scaling*, ed. W. E. Saris and I. N. Gallhofer, 137–57. London: Macmillan.

Van Schuur, W. H. 1989. "Unfolding the German Political Parties." In *Advances in Psychology*, vol. 60, *New Developments in Psychological Choice Modeling*, ed. G. De Soete, H. Feger, and K. C. Klauer, 259–90. Amsterdam: North-Holland.

APPENDIX: A SMALL NUMERICAL EXAMPLE FOR THE UNFOLDING SCALE ABCD

Response Pattern					Number of Errors											
					Unfolding Scale				Other Permutations							
A	B	C	D	Frequency	ABC	ABD	ACD	BCD	BAC	ACB	BAD	ADB	CAD	ADC	CBD	BDC
1	1	1	1	70												
1	1	1	0	240								240		240		240
1	1	0	1	40			40	40		40						
1	0	1	1	20	20	20									20	
0	1	1	1	24					24		24		24			
1	1	0	0	160						160		160				
1	0	1	0	60	60									60		
1	0	0	1	10		10	10									
0	1	1	0	16					16							16
0	1	0	1	14				14			14					
0	0	1	1	4									4		4	
1	0	0	0	168												
0	1	0	0	48												
0	0	1	0	8												
0	0	0	1	28												
0	0	0	0	90												
Frequency																
768	612	442	210	1000	80	30	50	54	40	200	38	400	28	300	24	256
Errors expected					132	63	90	72	63	208	30	371	22	268	36	214
$H(ijk)$					.39	.52	.44	.25	.37	.04	−.27	−.08	−.27	−.12	.37	−.20

Notes: The number of errors expected in the unfolding scale for triple ABC was calculated as: .768 * (1000 − 612) * .442 = 124.51. H(ABC) was calculated as 1 − 80/132 = 0.39.

Two triples are unique: ABD ($H_{ABD} > 0$ and H_{BAD}, $H_{ADB} < 0$), and ACD. Triple ABD is selected because the number of subjects who give the positive response to the 111, the 110, or the 011 pattern is higher than for triple ACD. Its $H(\text{ABD}) = 0.52$.

Item C is admissible in the second and third position in triple ABD to make either scale ACBD or scale ABCD. The last scale has a higher H-value (0.40) than the first one (0.23), as the reader may verify. Hence, scale ABCD is the final scale. Homogeneity coefficients for each item and for the whole scale are computed below.

E(o)A: 160 (80 + 30 + 50)	E(e)A: 285 (132 + 63 + 90)	H(A) = .44
E(o)B: 164 (80 + 30 + 54)	E(e)B: 267 (132 + 63 + 72)	H(B) = .39
E(o)C: 184 (80 + 50 + 54)	E(e)C: 294 (132 + 90 + 72)	H(C) = .37
E(o)D: 134 (30 + 50 + 54)	E(e)D: 225 (63 + 90 + 72)	H(D) = .40

The scalability of the whole scale has the following coefficients: E(o): 214 (80 + 30 + 50 + 54); E(e): 357 (132 + 63 + 90 + 72); H = .40.