KM 42 (1993)
pg.143-160

Latent Class Models for Monotone and Nonmonotone Dichotomous Items

Anton K. Formann

Department of Psychology, University of Vienna

A variety of scaling models of the latent class type is known for monotone dichotomous items. Both these models and their analogues for nonmonotone items are sketched in this paper. They are probabilistic in nature, allowing response errors for the omissions as well as for the intrusions, and can further be extended by providing for intrinsically unscalable respondents whose response behavior is not governed by the scaling model. All these models can be understood as simply restricted latent class analysis, so that the estimation and identifiability of the parameters (class sizes and item latent probabilities) as well as the goodness-of-fit tests are free of problems. The applicability of these models is demonstrated on two sets of data concerning the attitude towards nuclear energy and the attitude towards car-use and environment, respectively. While the five items of the first data could be substantiated to be, at least in part, nonmonotone, the ten items of the second scale seem to be monotone, even if they were expected to be nonmonotone; and only five of them can be put together to form a scale in the sense of the latent distance model.

Requests for reprints should be sent to Anton K. Formann, Department of Psychology, University of Vienna, Liebiggasse 5, A–1010 Wien (Vienna), Austria.

The theoretical part of this contribution and the first example mainly duplicate results reported by the same author in an earlier paper (Psychometrika, 1988).

The author wishes to thank Herbert Hoijtink and an anonymous referee for their helpful comments on an earlier draft of the present paper.

The Unconstrained Latent Class Model for Dichotomous Data

Let every person P_a $(a = 1, \ldots, N)$ of a sample of size N respond to n dichotomous items I_i $(i = 1, \ldots, n)$. The reaction of each subject P_a to each item I_i is denoted by x_{ai}, and takes the values "1" (positive response) and "0" (negative response), respectively. By means of three assumptions, homogeneous groups of persons who differ from each other by their reaction tendencies with respect to the items are sought:

1. The population, from which the observed sample is drawn, consists of T latent classes $C_1, \ldots, C_t, \ldots, C_T$ of relative sizes π_t, with
$$\sum_t \pi_t = 1. \tag{1}$$
2. For each item I_i, every class C_t has a specific probability of positive responses, $x_{ai} = 1$,
$$p(x_{ai} = 1|C_t) = p_{i|t}. \tag{2}$$
3. Within each class the responses are stochastically independent, so that the probability of all responses $\boldsymbol{x}_a = (x_{a1}, \ldots, x_{an})$ of subject P_a from class C_t can be written as the product of the probabilities of that person's individual responses, that is,
$$p(\boldsymbol{x}_a|C_t) = \prod_i p_{i|t}^{x_{ai}}(1 - p_{i|t})^{1-x_{ai}}. \tag{3}$$
By weighting with the class sizes and summing over all classes, the unconditional probability of the response vector $\boldsymbol{x}_a$ results,
$$p(\boldsymbol{x}_a) = \sum_t \pi_t \prod_i p_{i|t}^{x_{ai}}(1 - p_{i|t})^{1-x_{ai}}. \tag{4}$$

Since there are n dichotomous items, there exist only $S = 2^n$ distinguishable response patterns, whose relative frequencies of occurrence, corresponding to the left-hand side of (4), can be obtained from the sample. Subsequently, the unknown π_t's and $p_{i|t}$'s can be estimated, for example, by the method of maximum likelihood (Clogg, 1977; Goodman, 1974; McHugh, 1956).

Latent Class Models for Scaling Monotone Items

Perfectly discriminating monotone items form the well-known Guttman (1950) scale which allows only a certain subset out of all possible response patterns. For example, for $k = 3$ items ordered according to ascending difficulty, the following 4 out of the $2^3 = 8$ patterns are admissible: (0,0,0), (1,0,0), (1,1,0), (1,1,1).

To weaken these unrealistically strong assumptions of the perfect Guttman scale, Goodman (1975) adds to these scalable types a further type of the unscalable which produces all possible response patterns. As was also shown by Goodman, this model can be interpreted as a restricted latent class model (Goodman, 1974): Each of the scalable types corresponds to a class whose item latent probabilities are restricted deterministically to 0 or 1, so that each of these classes produces only his response pattern, and the unscalable class possesses unrestricted item latent probabilities. The unscalable class can be understood to be a rest category of persons not obeying the response laws met by the deterministic model; preferably it should be small.

A somewhat different possibility to relax the strong assumption of items with perfect discriminatory power is realized by the so-called latent distance models (see Lazarsfeld & Henry, 1968, chap. 5) providing item-specific unknown probabilities of errors for the omissions (instead of the positive, the negative response is observed) and the intrusions (instead of the negative, the positive response occurs). Combining the latent distance models with the Goodman model with intrinsically unscalable respondents leads to more flexible variants which in turn are also specifically restricted latent class models (Dayton & Macready, 1980).

The Deterministic Scaling Model for Dichotomous Point Items

Assuming perfectly discriminating point items, their trace lines are discontinuous functions with two discontinuities: Each person P_a with attitudinal parameter β_a never agrees with item I_i, if β_a lies under the first cutting point, δ_{1i}, or if β_a lies over the second cutting point, δ_{2i}, of that item; on the other hand, if β_a lies between δ_{1i} and δ_{2i}, the probability of agreement with item I_i is 1. Therefore, the probability of the positive response, $p(x_{ai} = 1|\beta_a)$, being dependent upon the latent parameter β_a, is for item I_i

$$p(x_{ai} = 1|\beta_a) = \begin{cases} 0 & \text{for} \quad \beta_a < \delta_{1i} \text{ and } \beta_a \geq \delta_{2i}, \\ 1 & \text{for} \quad \delta_{1i} \leq \beta_a < \delta_{2i}. \end{cases} \tag{5}$$

If the item characteristic (IC-)curves do not overlap, neither the order of the items nor the order of the subjects can be determined: Independent of the order of the items, the set of scalable response patterns remains the same; it comprises those patterns containing exactly one positive response. For those persons showing exclusively negative responses, their position on the underlying continuum does not coincide with the position of any item; thus, they cannot be located on the latent scale and have to be considered unscalable. If, on the contrary, adjacent pairs of IC-curves of ordered items are assumed to overlap, the ordering of the items becomes relevant since rearranging the order of the items actually changes the subset of admissible patterns.

In the deterministic case (errorless data), the order of the items can be determined by simply interchanging the items and the response patterns. If the data can be arranged so that the 1's create a parallelogram structure, the items form a perfect scale. When columns and rows of the data matrix are arranged optimally, any departure from the parallelogram structure indicates a departure from the perfect scale. The order of the items and the order of the subjects on the underlying continuum is given by the arrangement of the items and of the response patterns generating the parallelogram (Torgerson, 1958, p.314, p.316).

Probabilistic Scaling Models for Dichotomous Point Items

Transferring Goodman's concept of the unscalable from the monotone to the nonmonotone items, to the scalable types (there are n in the case of non-overlapping items, and $2n - 1$ in the case of pair-wise overlapping items) a further type must be added providing for those respondents who do not conform to the deterministic model. Applying the concept of latent distance models to nonmonotone items takes response errors into account for the scalable respondents (Formann, 1988).

Response-error model RE0. For point items, the most general form of IC-curves (see Figure 1) assumes three probabilities for positive responses per item I_i, a high

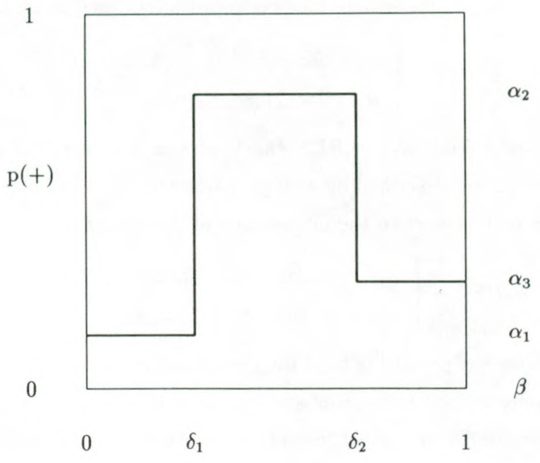

Figure 1: Trace Line of a Dichotomous Point Item with Three Levels of Agreement.

probability α_{2i} of agreement for persons in the middle of the latent dimension, and two low probabilities α_{1i} and α_{3i} for persons with extremely positive and extremely negative attitudes:

$$p(x_{ai} = 1|\beta_a) = \begin{cases} \alpha_{1i} & \text{for} \quad \beta_a < \delta_{1i}, \\ \alpha_{2i} & \text{for} \quad \delta_{1i} \leq \beta_a < \delta_{2i}, \\ \alpha_{3i} & \text{for} \quad \beta_a \geq \delta_{2i}. \end{cases} \tag{6}$$

Thereby it is assumed that the persons' distribution over the latent trait is uniform in the interval [0,1], so that the cutting points δ_{1i} and δ_{2i} of the IC–curves and the persons' positions β_a on the latent continuum can be interpreted as probabilities: δ_{1i} and δ_{2i} fix the proportions with which persons fall into the three sections of the latent trait that are formed by each one item, and β_a gives the relative position of subject P_a within the population in the sense of percentiles.

Special cases of (6) arise by equating some of the probabilities describing the IC–curves; cf. Formann (1988), formulas (7) to (11).

Response-error model RE2. The IC-curves are assumed to be specific for each item, but this model provides two different levels of agreement instead of three, the one for persons with high agreement, the second for persons with low agreement:

$$p(x_{ai}=1|\beta_a)=\begin{cases}\alpha_{1i} & \text{for} \quad \beta_a<\delta_{1i} \text{ and } \beta_a \geq \delta_{2i}, \\ \alpha_{2i} & \text{for} \quad \delta_{1i} \leq \beta_a < \delta_{2i}.\end{cases} \tag{7}$$

Response-error model RE4. As in RE2, the IC-curves are specific for each item, but symmetric. So, they are described by a single parameter α_{1i} each, standing for the response errors both with respect to the intrusions and the omissions:

$$p(x_{ai}=1|\beta_a)=\begin{cases}\alpha_{1i} & \text{for} \quad \beta_a<\delta_{1i} \text{ and } \beta_a \geq \delta_{2i}, \\ 1-\alpha_{1i} & \text{for} \quad \delta_{1i} \leq \beta_a < \delta_{2i}.\end{cases} \tag{8}$$

Even if this model looks less promising from the conceptual point of view because intrusions and omissions are caused by psychologically distinct processes, so that equating their probabilities can hardly be argued, from the statistical point of view it may be of interest due to its parsimony.

Each of the response-error models can be defined for varying sets of scalable response patterns, generated by nonoverlapping or overlapping trace lines of 2,3,... items, and can be further extended to incorporate unscalable respondents. Assuming local stochastic independence of the answers of each person, all resulting models can be seen as restricted latent class models in Goodman's (1974) sense: Whereas the class sizes π_t, whose number equals the number of scalable types (plus 1, when a group of unscalable respondents is assumed to exist), remain unconstrained, certain of the item latent probabilities $p_{i|t}$ have to be set equal to each other. Table 1 contrasts for $n = 3$ items the latent probabilities of the unrestricted latent class model to those of the response-error model RE0 under the additional consideration of unscalable respondents and assuming overlapping IC-curves for pairs of adjacent items.

Since all point item models mentioned above are specifically restricted latent class models, they are statistically testable by means of standard goodness-of-fit tests (Pearson and likelihood-ratio chi-squared statistics) in the usual manner. Thereby the degrees of freedom are given by the difference of the number of all possible response patterns minus 1, that is 2^n-1, and the number of independent parameters to be estimated, that is the number of classes minus 1, $T-1$, plus the number of parameters describing the

TABLE 1
Latent Class Parameters for the Unrestricted Model (A), and the Response-Error Model RE0 with a Class of Unscalable Respondents for Three Pair-Wise Overlapping Items (B).

CLASS	CLASS SIZE	A: ITEM LATENT PROBABILITIES			B: SCALABLE RESPONSE PATTERN	B: ITEM LATENT PROBABILITIES		
		I_1	I_2	I_3		I_1	I_2	I_3
1	π_1	$p_{1\|1}$	$p_{2\|1}$	$p_{3\|1}$	100	α_{21}	α_{12}	α_{13}
2	π_2	$p_{1\|2}$	$p_{2\|2}$	$p_{3\|2}$	110	α_{21}	α_{22}	α_{13}
3	π_3	$p_{1\|3}$	$p_{2\|3}$	$p_{3\|3}$	010	α_{31}	α_{22}	α_{13}
4	π_4	$p_{1\|4}$	$p_{2\|4}$	$p_{3\|4}$	011	α_{31}	α_{22}	α_{23}
5	π_5	$p_{1\|5}$	$p_{2\|5}$	$p_{3\|5}$	001	α_{31}	α_{32}	α_{23}
6	π_6	$p_{1\|6}$	$p_{2\|6}$	$p_{3\|6}$	unscalable	$p_{1\|u}$	$p_{2\|u}$	$p_{3\|u}$

trace lines. The restrictions used in the latent class models to realize the assumptions met here consist of constant (fixed to 0 or 1) or equated item latent probabilities, so that the ML parameter estimation can be performed by means of available computer programs, for example by means of that developed by Clogg (1977). This program provides for a test of local identifiability (Goodman, 1974; McHugh, 1956) which is all the more important as not all models must be identifiable (e.g., the response-error model RE0 is not identifiable, in principle).

Within the framework of latent class analysis, scaling each respondent consists in computing the conditional (or posterior) class membership probabilities $p(C_t|\boldsymbol{x}_a)$ of that response pattern he has shown:

$$p(C_t|\boldsymbol{x}_a) = \pi_t p(\boldsymbol{x}_a|C_t)/p(\boldsymbol{x}_a). \tag{9}$$

In unconstrained latent class analysis, $p(C_t|\boldsymbol{x}_a)$ simply gives the probability with which each pattern belongs to each class; in scaling models of the latent class type, where the classes correspond to ideal response patterns, the class membership probabilities quantify the affinity of each pattern to the ideal ones. This gives the possibility to arrange the non-ideal response patterns on the latent scale fixed by the ideal response patterns, and is of special interest for the unscalable patterns in case that an unscalable class is assumed to exist: Without having at hand this information for the unscalable patterns, nothing would be known concerning their relation to the scalable patterns; contrary to this, for the scalable patterns their order on the latent scale is determined by the order of the items.

The strategy of class assignment being optimal with respect to classification errors assigns each response pattern $\boldsymbol{x}_a$ deterministically to that class C_a^* with the highest class membership probability,

$$p(C_a^*|\boldsymbol{x}_a) = \max_t \; p(C_t|\boldsymbol{x}_a). \tag{10}$$

For the possibility to assign scale values to the patterns, see Formann (1988, p.54).

Finally, the problem is to be mentioned that the order of the items must be given to know the scalable response patterns. This leads to the same practical complication as in the presence of monotone items if one wishes to apply those latent distance models for which the latent order of the items is not necessarily identical with the manifest order according to their item marginals: For each set of empirical data, the order of the items must be determined prior to the parameter estimation by means of one of the latent class models for nonmonotone items. In case that this is not possible a priori, that means, based on theoretical grounds, several methods are available, in principle, to find out the correct order, or, at least, an approximation to it. Some of these methods will be sketched in the following. First, by interchanging the columns (items) and rows (response patterns) of the data matrix one can try to approximate best the parallelogram structure described for the deterministic (Guttman) approach of point items; if, after rearranging, the most frequent response patterns create a parallelogram, then that order of the items can be used as the initial configuration for performing analyses by means of latent class models. Second, one can compute the intercorrelations of the items; it can be expected that neighboring items correlate higher than others, and that the correlation decreases with increasing distance between the items. (Note, however, that this plausible claim cannot be proven without specifying the IC–curves; but even within that class of models described here, its proof is possible for simple special cases only.) Third, the cross-tabulation of positive versus negative responses to each item with an external categorical criterion reflecting the attitude under consideration – if available – gives an impresssion of the IC-curves of all items; that category of the external criterion where each item has its highest probability of agreement can be used to estimate the position of each item on the underlying attitudinal continuum. Fourth, the correct order of the items can be ascertained by comparing the goodness-of-fit statistics for varying orders of the items. Whereas the first three methods bear no computational problems, the fourth method is practicable only if the order of 2 or 3 items was unclear; otherwise, the computational effort would become prohibitive.

Empirical Example 1: Attitude Towards Nuclear Energy

The results to be presented in the following summarize those given by Formann (1988) and refer to a sample of $N = 600$ persons and $n = 5$ items on attitude towards nuclear energy. To get the presumable order of the items, their intercorrelations and probabilities for positive responses dependent upon the general attitude towards nuclear energy, which was available as a further, external criterion, were computed. Both methods led to the same result, whereby positive responses to item I_1 indicate the most favorable attitude towards nuclear energy, and positive responses to item I_5 the most disapproving attitude. Since the formulation of the items I_1 and I_5 as well as their cross classification against the general attitude towards nuclear energy suggest monotonicity of their IC-curves, the five items under consideration can be seen to be a mixture of two (inversely scored) monotone and three nonmonotone items.

Several analyses assuming different patterns to be scalable as well as different response error models with and without an additional class of unscalable respondents were performed. Finally, they led to the conclusion that a rather complex model (RE2/4, that is RE4 for the items I_1 and I_5, and RE2 for the items I_2, I_3, and I_4, with a class of unscalables) is not to be rejected if the following ideal response patterns are assumed: (1,0,0,0,0), (1,1,0,0,0), (1,1,1,0,0), (0,1,1,0,0), (0,1,1,1,0), (0,0,1,1,1), (0,0,0,1,1), and (0,0,0,0,1). For this nine-classes model (8 scalable classes plus 1 unscalable), the parameter estimates together with the goodness-of-fit statistics are given in Table 2.

Most of the respondents are unscalable ($\pi_9 = .424$). The sizes of the classes corresponding to the scalable patterns differ considerably from each other; whereas π_1 – for the perfect pattern (1,0,0,0,0) – and π_2 – for (1,1,0,0,0) – are very small, considerable

TABLE 2

Attitude Towards Nuclear Energy – Goodness-of-Fit Tests, Class Sizes and Item Latent Probabilities for the Response-Error Model RE2/4 with a Class of Unscalable Respondents.

CLASS	SIZE	ITEM LATENT PROBABILITIES				
		I_1	I_2	I_3	I_4	I_5
1	.003	.936	.267	.014	.237	.006
2	.002	.936	.503	.014	.237	.006
3	.091	.936	.503	.807	.237	.006
4	.083	.064	.503	.807	.237	.006
5	.020	.064	.503	.807	.997	.006
6	.317	.064	.267	.807	.997	.994
7	.033	.064	.267	.014	.997	.994
8	.028	.064	.267	.014	.237	.994
9	.424	.476	.629	.974	.929	.286
$X^2 = 10.79$, $L^2 = 11.53$, $df = 10$, $\chi^2_{95} = 18.31$						

Note: X^2 = Pearson's chi-squared statistic; L^2 = likelihood ratio test.

portions of the sample belong to the patterns (1,1,1,0,0), (0,1,1,0,0) and especially to (0,0,1,1,1) whose class size π_6 is .317. The cutting points of the items result from the class sizes. Their relations depend upon the response patterns which were defined to be scalable, so that no general formulas can be given. For more details, see the two hypothetical examples in Formann (1988, pp. 52–53) and the calculations referring to the present example being described there (p. 59). In their correct order on the latent continuum, here the cutting points are $\delta_{11} = 0$, $\delta_{12} = .003$, $\delta_{13} = .005$, $\delta_{21} = .096$, $\delta_{14} = .179$, $\delta_{15} = \delta_{22} = .199$, $\delta_{23} = .516$, $\delta_{24} = .549$, $\delta_{25} = .577$. Thus, for the ranges $R_i = \delta_{2i} - \delta_{1i}$ of high probability of agreement per item, one obtains $R_1 = .096$, $R_2 = .196$, $R_3 = .511$, $R_4 = .370$, $R_5 = .378$ showing that with respect to this criterion of discriminatory power, the items I_2, I_4, and I_5 can be classified as being good, while the items I_1 and I_3 must be classified as being poor discriminators. (For example, the subsample of the scalables is divided by item I_2 at the rate .196: .381 $\approx$ 1:2, but by item I_3 at the rate .066: .511 $\approx$ 1:8; that is, nearly all scalable respondents show the same – high – probability for positive reactions at item I_3, while this is not the case for item I_2.) The latent probabilities of each item may be used as the second criterion for discriminatory power: The greater α_{1i} and α_{3i}, respectively, the greater are the response errors with respect to the intrusions; the smaller α_{2i}, the greater are the response errors with respect to the omissions; and, ideally, $\alpha_{1i} = \alpha_{3i} = 0$ and $\alpha_{2i} = 1$, which holds for the deterministic point items. With respect to this second criterion, for the five items under consideration a somewhat different picture emerges as compared to the first criterion, since now only for item I_2 the discrepancy between the low and the high probabilities of agreement is unsatisfying ($\alpha_{12} = .267$, $\alpha_{22} = .503$). On the other hand, the items I_1 and I_5 are nearly perfect ($\alpha_{11} = .064$, $\alpha_{21} = 1 - \alpha_{11} = .936$; $\alpha_{15} = .006$, $\alpha_{25} = 1 - \alpha_{15} = .994$), and the items I_3 and I_4 discriminate quite efficiently.

Since more than 40 percent of the sample belong to the unscalable class, for a relatively large portion of the response patterns this class is the most probable, too. Thereby, using (9), even one of the scalable patterns will be assigned to the unscalable class, namely the pattern (0,1,1,1,0). Inspecting the class membership probabilities further, it becomes evident that no response patterns will be assigned to the classes C_1 and C_2 at all, and that only the scalable patterns (1,1,1,0,0), (0,1,1,0,0), (0,0,1,1,1), and (0,0,0,0,1) belong to the corresponding scalable class; the remaining scalable patterns (1,0,0,0,0), and (1,1,0,0,0) will be assigned to the "false" scalable class C_3 and the pattern (0,0,0,1,1) will be assigned to the "false" class C_6. But contrary to this, in the case of some of the response patterns not being conform to one of the scalable classes, they will be assigned to one of these latter.

Concerning the measurability of the attitude towards nuclear energy, it can be concluded that the five items under study form a onedimensional scale, however, for less than 60% of the sample only; the remaining subjects do not fit into that scale. Even if this finding looks somewhat unsatisfactory, it seems typical of scaling models providing for unscalable respondents; cf. the results reported in Goodman (1975) what concerns the application of the quasi-independence model to monotone items.

Empirical Example 2: Attitude Towards Car-Use and Environment

In the Netherlands, 10 statements concerning car-use and environment were presented to two samples of 300 persons each at two occasions, the first one before and the second one after a campaign. The dichotomously scored items were suspected to be nonmonotone.

TABLE 3

Attitude Towards Car-Use and Environment – Response Patterns, $\boldsymbol{a}_s$, and Their Observed Frequencies, n_s, Before and After the Campaign, for the Items I_1, I_2, I_5, I_7, and I_8.

$\boldsymbol{a}_s$ 12578	n_s PRE	n_s POST	$\boldsymbol{a}_s$ 12578	n_s PRE	n_s POST	$\boldsymbol{a}_s$ 12578	n_s PRE	n_s POST	$\boldsymbol{a}_s$ 12578	n_s PRE	n_s POST
11111	51	71	10111	66	39	01111	9	12	00111	7	4
11110	8	22	10110	7	9	01110	2	2	00110	1	–
11101	21	33	10101	14	5	01101	7	5	00101	4	1
11100	15	15	10100	4	1	01100	4	4	00100	1	1
11011	11	8	10011	3	4	01011	3	3	00011	–	–
11010	7	6	10010	2	–	01010	3	2	00010	–	–
11001	20	23	10001	2	2	01001	2	2	00001	–	1
11000	11	17	10000	1	1	01000	14	7	00000	–	–

- Item 1: Car use cannot be abandoned. Some pressure on the environment has to be accepted.
- Item 2: A cleaner environment demands for sacrifices like a decreasing car usage.
- Item 5: It is better to deal with other forms of environmental pollution then car driving.
- Item 7: Technically adapted cars do not constitute an environmental threat.
- Item 8: Considering the environmental problems, everybody should decide for themselves how often to use the car.

Therefore, it was intended first to apply those types of latent class models, which were used for analyzing the attitude towards nuclear energy, to the pre-test data on car-use and environment, and second, to validate the solution found for the pre-test data on the post-test data; finally, the changes caused by the campaign should be inferred by applying simultaneous latent class analysis (Clogg & Goodman, 1984) to the pre- and post-test data.

Because of the greater number of items, the smaller sample size, and the lack of additional external criteria reflecting general attitude towards car-use and environment, the data on car-use and environment are harder to analyse than those on nuclear energy: As a consequence of the lack of additional external criteria, the presumable order of the items can be derived only from their intercorrelations and from the frequency distribution of the response patterns (parallelogram structure of patterns having been observed frequently); however, the latter one is more difficult for 10 items and 300 persons than for 5 items and 600 persons. In addition, the asymptotic χ^2-distribution of the goodness-of-fit statistics becomes doubtful for sparse frequencies so that the fit of a certain model assuming a certain order of the items cannot be judged correctly.

Taking all this into consideration, the strategy has been changed: Instead of analyzing all 10 items of the pre-test data together (with the further intention to select inappropriate items later, if necessary), some overlapping subsets of items were defined, for example, one of them containing the items I_1 to I_5, another one containing I_6 to I_{10}, a third one containing I_4 to I_7, and so on. For each subset its response patterns' frequencies were counted, and then used to fix the order of the items within that subset by applying the parallelogram criterion; after determining the presumable order within each subset, to each one of them latent class analysis was applied in order to assess the fit. As a result of this tentative search strategy, finally for the five items I_1, I_2, I_5, I_7, and I_8, see Table 3, and the 4 ideal response patterns (0,1,0,0,0), (1,1,0,0,0), (1,1,1,1,1), and (1,0,1,1,1), the fit was found to be very good if item specific response errors are allowed for both the intrusions and the omissions. As can be seen from the solution presented in Table 4 (H_0, t_1), in part the response errors of the intrusions are considerable (items I_1, I_5, and I_8), while they are rather small for the omissions. As can bee seen, too, the class sizes had to be restricted in order to get the model identifiable: The first three classes have the same size of about 20 % of the sample, the fourth class comprises about 40 % of the sample. (This type of restrictions is an alternative to restrictions of the latent probabilities for the first and the last item; cf. the solution described for the data on nuclear energy, Table 2). That none of the remaining items I_3, I_4, I_6, I_9, and I_{10} do fit into this scale will be shown later.

If the same model is applied to the post-test data, its fit is also very good, however, revealing slightly different parameter estimates as compared with those of the pre-test data; see Table 4 (H_0, t_2). To find out which changes must be rated to be relevant, simultaneous analyses of both data sets were performed assuming a) no changes (H_1), b) changes of the class sizes alone (H_2), and c) changes of the item latent probabilities alone (H_3). According to the results of the analyses summarized in Table 4 again, the hypothesis of no changes at all (H_1) must be discarded, the other two hypotheses can be maintained.

This can be concluded from the comparison of the three hypotheses with that one

TABLE 4

Attitude Towards Car-Use and Environment – Parameter Estimates and Goodness-of-Fit Tests for Some Hypotheses [Ideal Response Patterns: (0,1,0,0,0), (1,1,0,0,0), (1,1,1,1,1), and (1,0,1,1,1)].

HYPOTHESIS		CLASS	CLASS SIZE	ITEM LATENT PROBABILITIES I_1	I_2	I_5	I_7	I_8	HYPOTHESES TESTS	
H_0	t_1	1	.197	.514	.941	.409	.292	.481	$X^2 = 21.403$	
		2	.197	.883	.941	.409	.292	.481	$L^2 = 22.721$	
		3	.197	.883	.941	.950	.801	.898	$df = 17$	$X^2 = 35.698$
		4	.408	.883	.171	.950	.801	.898	$\chi^2_{95} = 27.59$	$L^2 = 39.265$
H_0	t_2	1	.210	.703	.942	.472	.225	.555	$X^2 = 14.295$	$df = 33$
		2	.210	.893	.942	.472	.225	.555	$L^2 = 16.544$	$\chi^2_{95} = 47.40$
		3	.210	.893	.942	.946	.884	.822	$df = 16$	
		4	.369	.893	.484	.946	.884	.822	$\chi^2_{95} = 26.30$	
H_1		1	.198	.603	.945	.432	.265	.504	$X^2 = 63.821$	
		2	.198	.888	.945	.432	.265	.504	$L^2 = 65.916$	
		3	.198	.888	.945	.945	.825	.865	$df = 44$	
		4	.406	.888	.340	.945	.825	.865	$\chi^2_{95} = 60.48$	
H_2			t_1 t_2							
		1	.198 .257	.619	.942	.473	.299	.527	$X^2 = 57.133$	
		2	.198 .257	.892	.942	.473	.299	.527	$L^2 = 55.366$	
		3	.198 .257	.892	.942	.955	.845	.876	$df = 43$	
		4	.405 .238	.892	.190	.955	.845	.876	$\chi^2_{95} = 59.30$	
H_3	t_1	1	.202	.519	.936	.415	.297	.485		
		2	.202	.884	.936	.415	.297	.485		
		3	.202	.884	.936	.954	.805	.901	$X^2 = 35.752$	
		4	.392	.884	.151	.954	.805	.901	$L^2 = 39.334$	
	t_2	1	.202	.701	.945	.461	.217	.549	$df = 34$	
		2	.202	.892	.945	.461	.217	.549	$\chi^2_{95} = 48.60$	
		3	.202	.892	.945	.943	.874	.821		
		4	.392	.892	.506	.943	.874	.821		
H_1 vs. H_0		$L^2 = 26.651$		$df = 11$		$\chi^2_{95} = 19.68$				
H_2 vs. H_0		$L^2 = 16.101$		$df = 10$		$\chi^2_{95} = 18.31$				
H_3 vs. H_0		$L^2 = 0.069$		$df = 1$		$\chi^2_{95} = 3.84$				

Notes: X^2 = Pearson's chi-squared statistic, L^2 = likelihood ratio test. Both the degrees of freedom (df) and the critical χ^2-values refer to the corrected values when collapsing those patterns having zero frequencies, cf. Table 3. The model selection was based on these corrected values.

t_1 = pre-test, t_2 = post-test.

H_0: changes of both the class sizes and the item latent probabilities.

H_1: no changes.

H_2: changes of the class sizes alone.

H_3: changes of the item latent probabilities alone.

which allows both changes of the class sizes and of the item latent probabilities (H_0) being equivalent to the separate analyses of both sets of data. The hypothesis of changes of the item latent probabilities alone (H_3) fits better, but the hypothesis of changes of the class sizes alone (H_2) seems to be more appealing: First, from the conceptual point of view, because H_2 represents a changing attitude in the population that is recorded with a constant instrument of measurement, whereas under H_3 that instrument itself changes, and second, from the statistical point of view, because H_2 is more parsimonious with respect to the number of parameters. It is interesting to note that assuming changes of the class sizes alone, they become rather massive as compared with the changes resulting from the separate analyses of both sets of data. That is, changes of the item latent probabilities that are not allowed under this hypothesis, obviously influence the extent to which changes in the class sizes result in case that solely such changes are allowed.

The changes in the class sizes caused by the campaign (and possibly also by the inseparable trend) find a simple interpretation after having had a look at the wording of the 5 statements, see Table 3, revealing that item I_2 should be inversely scored: Then, the positive answer at all 5 items indicates positive attitude towards car-use, and the order of the items is not I_2, I_1, (I_5, I_7, I_8) – the parantheses indicate that I_5, I_7, and I_8 take the same position –, but I_1, (I_5, I_7, I_8), and, inversely scored, I_2; the 5 monotone items I_1, I_5, I_7, I_8, and inversely scored, I_2, form a scale whose levels correspond to the ideal patterns (0,0,0,0,0), (1,0,0,0,0), (1,1,1,1,0), and (1,1,1,1,1), so that the solution based on the assumption of nonmonotone items is equivalent to Lazarsfeld's well-known latent distance model for monotone items. The effect of the campaign goes towards the intended direction in that those classes showing less positive answers concerning car-use become greater.

After having identified that five-items scale, one can investigate whether the remaining items conform to that scale or not. This can be done by separately analyzing each one of the remaining items together with those 5 items forming the scale, whereby the latent probabilities of the 5 items belonging to the scale are restricted according to the latent distance model for monotone items, and the latent probabilities of the additional item are unrestricted. The degree to which the additional item fits into the scale can

TABLE 5

Attitude Towards Car-Use and Environment – Goodness-of-Fit Tests and Estimated IC-Curves of Item I_6 as well as of the Inversely Scored Items I_3, I_4, I_9, and I_{10}, when Each One of Them is Analyzed Together with the Scale-Items I_1, I_5, I_7, I_8, and, Inversely Scored, I_2.

CLASS	CORRESPONDING IDEAL PATTERN OF THE SCALE	LATENT PROBABILITIES OF ITEM I_3	I_4	I_6	I_9	I_{10}
1	(0,0,0,0,0)	.163	.376	.039	.565	.312
2	(1,0,0,0,0)	1.000	.341	.045	.843	.484
3	(1,1,1,1,0)	.801	.684	.412	.792	.802
4	(1,1,1,1,1)	.969	.802	.695	.945	.839
FIT TESTS	X^2	59.189	58.934	42.563	39.149	54.207
	L^2	65.592	62.958	46.436	41.021	61.963
	df	30	33	31	34	32
	χ^2_{95}	43.77	47.40	44.99	48.60	46.19

- Item 3: The environmental problem justifies a tax burden on car driving so high that people quit using a car.
- Item 4: Putting a somewhat higher tax burden on car driving is a step in the direction of a healthier environment.
- Item 6: Instead of environmental protection measures with respect to car use, the road system should be extended.
- Item 9: People who keep driving a car, are not concerned with the future of our environment.
- Item 10: Car users should have to pay taxes per mile driven.

Note: Concerning the goodness-of-fit tests, see the remarks in Table 4.

be seen from the goodness-of-fit statistics, and the trace line of the additional item is estimated by its latent probabilities resulting for the classes being defined in terms of the latent distance model. So, information becomes available for each additional item concerning its (non-)monotonicity.

Applying this procedure to item I_6 as well as to the inversely scored items I_3, I_4, I_9, and I_{10} (pre–test data) turned out that none of them can be or needs to be included into the scale. For the items I_3, I_4, I_6, and I_{10}, this follows from the significance of the goodness-of-fit tests, and for the inversely scored item I_9, for which the fit is still acceptable, it follows from the estimated trace line: It is approximately monotone, but nearly the same as compared with that of item I_1. Because the trace lines of item I_6 and for the inversely scored items I_3, I_4, and I_{10} also seem to be approximately monotone, see Table 5, (Strictly speaking, they may not be interpreted because of the significant goodness-of-fit statistics.), perhaps the most striking result of the latent class analyses of the data on car-use and environment is, that monotonicity of all 10 items can be supposed after an appropriate scoring of the answers.

Final Remarks

The analyses of the first data (attitude towards nuclear energy) and their results showed the usefulness of latent class models for nonmonotone items. That the second data (attitude towards car-use and environment) unexpectedly were not conform to the assumption of nonmonotonicity, was at first sight disappointing. But, on the other hand, this insigth led to an impressive demonstration of the flexibility and generality of latent class analysis: Even if starting from false expectations, in the end a specific scaling model with sufficient goodness-of-fit could be identified for a five-items subset out of ten items, giving at least a ranking of the (monotone) items and of all the classes generated by them. (Note that in contrast to the first data, in case of the data on car-use and environment there is no need for an unscalable class.) In addition, this scaling model could be rediscovered for the post-test sample, so that simple hypotheses with respect to change due to the campaign became testable by means of simultaneous latent class analysis of the pre- and post-test data. So, the data on attitude towards car-use and environment also may serve as a nice example of assessing change in survey data using latent class analysis.

References

Clogg, C.C. (1977). Unrestricted and restricted maximum likelihood latent structure analysis: A manual for users. (Working paper 1977-09). University Park, PA: The Pennsylvania State University, Population Issues Research Center.

Clogg, C.C. & Goodman, L.A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association, 79,* 762-771.

Dayton, C.M. & Macready, G.B. (1980). A scaling model with response errors and intrinsically unscalable respondents. *Psychometrika, 45,* 343-356.

Formann, A.K. (1988). Latent class models for nonmonotone dichotomous items. *Psychometrika, 53,* 45-62.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61,* 215-231.

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, L.Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star & J.A. Clausen (Eds.), *Studies in social psychology in World War II,* Vol. IV: *Measurement and prediction.* Princeton: Princeton University Press.

Lazarsfeld, P.F. & Henry, N.W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

McHugh, R.B. (1956). Efficient estimation and local identifiability in latent class analysis. *Psychometrika, 21,* 331-347.

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