Chapter 18

Analysis of Student Problem Behaviors with Latent Trait, Latent Class, and Related Probit Mixture Models

John S. Uebersax

Department of Public Health Sciences
Wake Forest University, USA

1. Introduction

Here we investigate the latent structure of student problem behavior (Jessor & Jessor, 1977). The questions considered include Is problem behavior unidimensional or multidimensional?, Is problem behavior better viewed as a continuous or discrete trait?, and Are there different subtypes or „syndromes“ of problem behavior? Answers to these questions may increase our understanding of problem behavior and suggest interventions.

Researchers often approach such data with a single analytic technique. For example, they might use factor analysis to investigate trait dimensionality (e.g., Donovan & Jessor, 1985), or latent class analysis to identify population subtypes. The choice of method often has limited theoretical basis. A better approach, it is suggested here, is to apply a variety of modeling techniques. Doing so increases the chance of discovering something useful. Beyond this, comparisons of alternative models may permit inferences that would not otherwise be possible.

2. Models Considered

Four latent structure methods are considered. Two methods are well known and described elsewhere in this volume. The first is the standard Latent Class (LC) model (Goodman, 1974; Lazarsfeld & Henry, 1968; see chapter 1, section 2.). The second is the Latent Trait model (Bock & Aitkin, 1981; Lord & Novick, 1968; Rasch 1960, 1980; see chapter 1, section 1.1). We follow the Bock and Aitkin formulation of the latent trait model in supposing an underlying normal latent trait distribution, and denote this the LT model.

We also consider two less familiar models, as follows:

Located Latent Class model. Located Latent Class (LLC) models, originating with Formann (1985, 1992), have been described by several authors (Dayton & Macready, 1988; Lindsay, Clogg & Grego, 1991; Rost, 1988; Uebersax, 1993). These are, in one sense, a special case of the usual LC model, formed by imposing certain constraints on conditional response probabilities. In another sense, they are a variation of the LT model, in which the latent trait is assumed discrete – i.e., each latent class corresponds to a certain point on a unidimensional continuum.

Latent Mixture model. The Latent Mixture (LM) model (Mislevy, 1984; Uebersax & Grove, 1993) is an extension of the Bock and Aitkin LT model. With the latter, the latent distribution is assumed normal, and, in the simplest case, unidimensional. However, with the

LM model, the latent distribution is assumed a unidimensional mixture of two or more normal distributions.

Note that both the LLC and LM models are „in between“ the LC and LT models, because they combine the concepts of case subtypes and an underlying latent continuum. Uebersax (1994a) delineated statistical relationships among the LC, LT, LLC, and LM models. It was noted that, given normal-ogive item response functions for the LT, LLC, and LM models, all four models are special cases of a general ordered-categorical data mixture model described by Everitt and Merette (1990) and Henkelman, Kay, and Bronskill (1990). The general model, in essence, posits a mixture of multivariate normal latent distributions, and located thresholds which determine item response levels. The general model reveals formal hierarchical nestings among the LC, LT, LLC, and LM models, which, in turn, allow certain statistical inferences to be drawn by comparing their fits to a set of data.

With normal-ogive response functions, the LLC and LM models are examples of what can be termed *probit latent class* models. This class of models appears promising, especially as a way to relax conditional independence assumptions. For recent developments in this area, see Uebersax (1996) and Qu, Tan, and Kutner (1996).

3. Data and Methods

The data analyzed were collected during the 1991-92 academic year as part of a study of student substance use. Anonymous surveys were given to approximately 10,000 middle- and high-school students. Here we examine data only from male students in the 11th and 12th grades, and only for those with complete responses over eight problem behavior items; the final sample contained 834 students – 241 identified as black (African descent), 495 identified as white (European descent), and 98 with other or unknown ethnicity.

Table 1 shows the behaviors and percentages of respondents who reported each as having occurred in the preceding year.

Behavior	*Rate**
Sent from classroom	42.0
Fighting	34.9
Destroyed property	34.5
Trouble with parents	34.7
Suspended from school	22.1
Trouble with police	26.0
Drinking and driving	28.1
Car accident	15.6

*Percent reporting one or more occurrence of behavior in preceding 12 months.

Table 1: Student problem behaviors and rates (N = 834)

Raw data were used to generate cross-classification frequencies, which served as input for latent structure analysis. LC models with from two to five classes were examined. Several types of LLC models were estimated – including both the basic version and versions with added constraints – with from two to five latent classes. A unidimensional LT model was estimated, as was a two-component LM model, which assumed equal variances for both components. PANMARK (van de Pol, Langeheine & de Jong, 1989) was used to estimate the standard LC model, and LLCA (Uebersax, 1994b) to estimate the LLC model. The LT and LM models were estimated with a FORTRAN program written by the author. The LT, LLC, and LM models used normal-ogive item response functions, with a separate slope (correlation) parameter for each item.

4. Overall Results

Results for the more noteworthy models are summarized in Table 2. C2 – C5 denote the LC models with from two to five classes, respectively.

Model	*Latent distribution*	*No. of pars.*	*df*	G^2	X^2	*AIC*	*BIC*
C1	Independence	8	247	910.7	5388.8	7929	7967
C2	2 classes	17	238	343.4	344.1	7380	7460
C3	3 classes	26	229	289.2	264.4	7343	7466
C4	4 classes	35	220	254.9	248.0	7327	7493
C5	5 classes	44	211	233.7	223.4	7324	7532
M1	1 normal	16	239	293.0	259.9	7327	7404
M2	2 normals	18	237	293.0*	259.9*	--	--
L2	2 located classes	17	238	343.4	344.1	7380	7460
L5'	5 located classes + equally spaced	20	235	291.1	256.6	7333	7428

* Upper-bound estimates; see text for explanation.

Table 2: Results of latent class and latent trait models applied to data

M1 denotes the LT model and M2 the two-component LM model. L2 denotes a two-class LLC model and L5' a five-class LLC model with the added restriction of equally-spaced latent classes – this restriction assumes a fixed interval between successive latent classes and facilitates estimation by reducing the number of estimated model parameters (Uebersax, 1993). Shown for each model are two goodness-of-fit indices, the likelihood ratio chi-square (G^2) and Pearson chi-square (X^2) statistics; of these two, G^2 is generally preferred, particularly since it permits statistical comparison of nested models.

Two parsimony indices, the Akaike Information Criterion (AIC) and Schwarz' Bayes information criterion (BIC) are also shown. It has been this investigator's experience that the BIC, which gives greater penalty to extra parameters, gives somewhat better results for these types of models. For computational details on the AIC and BIC, see Sclove (1988).

5. Unidimensionality

We first consider whether adolescent problem behavior is unidimensional. Models M1 and L5', which both assume a unidimensional latent trait, are the most parsimonious by the BIC, suggesting the latent trait may in fact be unidimensional. Statistical comparison of models C5 and L5' provides additional information. The most important difference between these two models is that whereas L5' restricts all classes to fall along a single dimension, C5, in essence, permits them to be located anywhere in a four-dimensional space (Uebersax, 1994a). L5' is thus a special case of, and nested within, C5. The difference G^2 for the L5' - C5 comparison is 291.1 - 233.7 = 57.4 on 24 df. Due to sparse data, this difference, not large, probably cannot be interpreted in a strict statistical sense, and the result is somewhat equivocal as regards unidimensionality. Overall, perhaps the best conclusion is that the data do appear to have a salient primary dimension, but it does not completely account for the structure.

With more advanced models one might improve fit of a unidimensional model. There may be „spurious“ positive or negative associations among pairs of items, conditional on latent trait level. The former may occur when two items are very similar or redundant. The latter may occur when two items tap alternative outlets or expressions of a common underlying drive or tendency, such that one or the other may occur, but usually not both. A LISREL analysis based on tetrachoric item correlations, for example, could account for such associations in terms of correlated measurement error.

In any case, one must be cautious about inferring from such results a concrete trait that resides in adolescents themselves – reifying the latent construct. One could, for example, argue that the primary dimension, if it does exist, is mainly a property of the items. Intuitively, and, it would appear, statistically, the items have characteristics of a weak Guttman ordering. For example, students who have trouble with police probably also have trouble with parents, but the converse is not true as a general rule. There is something almost tautological about the apparent dimension here; one could question whether it says much beyond the obvious „some behaviors are more serious than others, and students with serious problems have usually also displayed milder problems.“

The dimension thus serves as a description, but not necessarily an explanation of problem behavior. It may be a common final pathway for a multitude of causal factors, and does not, by itself, indicate that all problem behaviors are manifestations of some fundamental personality characteristic (Gottfredson & Hirschi, 1990).

6. Continuous or Discrete Trait?

A second question is whether the latent trait is continuous or discrete. This is a common question with latent structure analysis, and often takes the form of a choice between the LT model or LC model as a way to represent data. Since neither of these models is nested within the other, direct comparison via the G^2 statistic is not possible. Some information is provided by comparing parsimony indices for the two models – in the present case, we observe that the LT model, M1, has a lower BIC value than any of the LC models, implying a continuous latent trait.

Of course, exclusive reliance on parsimony indices is questionable, especially since different indices may lead to different conclusions. It is of some interest, therefore, that, by supplementing the LT and LC models with the LLC and LM models, we can make formal statistical inferences about the latent distribution. A k-class LLC model is strictly nested within a k-component LM model. Specifically, the LLC model is created by restricting within-component variance of the LM model to 0. Comparison of a k-class LLC and k-class LM model, with k degrees of freedom, tests whether there is within-group variation on the latent trait, and, effectively, whether the trait is continuous or discrete.

Here there is a slight problem with the two-component LM model M2 – but, as we shall see, it leads to an interesting and potentially useful result. Specifically, M2 is not identified for these data; the problem appears related to what is termed a „Heywood case“ in factor analysis. However, we know that a two-component LM model must fit at least as well as M1, the one-component model. Therefore, as an upper-bound estimate for the fit of M2, we may take the G^2 value for M1.

Using this value, the difference G^2 for the L2 - M2 comparison is estimated as 343.4 – 293.0 = 50.4 on 238 – 237 = 1 df (note we use the df appropriate for M2, not for M1), which seems large enough to imply statistical significance. Moreover, this is a lower-bound estimate for the difference G^2. We thus conclude that M2 fits better than L2 and infer the superiority of a continuous distribution model.

We could apply similar logic for models with $k > 2$ component distributions and located classes; for instance, the G^2 for a two-component LM model gives an upper bound for the fit of a three-component LM model, and could be compared to the fit of a three-class LLC model. This principle may be especially helpful given that LM models become increasingly difficult to identify and estimate as the number of components increases.

Further, the LC and LLC models are equivalent when there are two latent classes, items are dichotomous, and the LLC model allows a separate slope parameter for each item. In the present case, L2 and C2 are equivalent, so that C2 is also nested within M2. One therefore obtains the same result and conclusion as above by comparison of C2 with M2, or, as here, C2 with M1. We observe (1) a statistical test of continuous versus discrete latent distribution is possible, and (2) in the simplest case it requires only a standard LT and standard two-class LC model – a potentially useful result and one possibly not previously reported in the literature.

7. Problem Behavior Subtypes

The results above suggest there is value in considering a unidimensional and continuous latent trait. However, this by no means precludes the utility of an LC model as an alternative way to explore and represent the data.

The fit and parsimony indices for models C2 – C5 provide no compelling evidence for the superiority of one over the others. The G^2 for C5 is low relative to df., but examination of the conditional response probabilities for the model suggests an overfit solution. Here we focus attention on the solution for C4 – more because it is substantively interesting than because of any statistical criterion.

Figure 1 shows the model's conditional probabilities. The latent classes are numbered to correspond to increasingly severe problem behavior: Class 1 is a low-problem group, Class 4 a high-problem group, and Classes 2 and 3 middle groups. The prevalence estimates for Classes 1 – 4 are .405, .273, .194, and .128, respectively.

Interpretation of Class 4 and Class 1 seems straightforward; there appears to be one type of student generally prone to problem behavior, and another type likely to have, or at least report, few problems. Perhaps more interesting is that there appear to be two qualitatively different middle groups. Class 3 students have a significantly higher probability of property damage, trouble with parents, and trouble with police; Class 2 students are significantly more likely to be sent from the classroom and suspended.

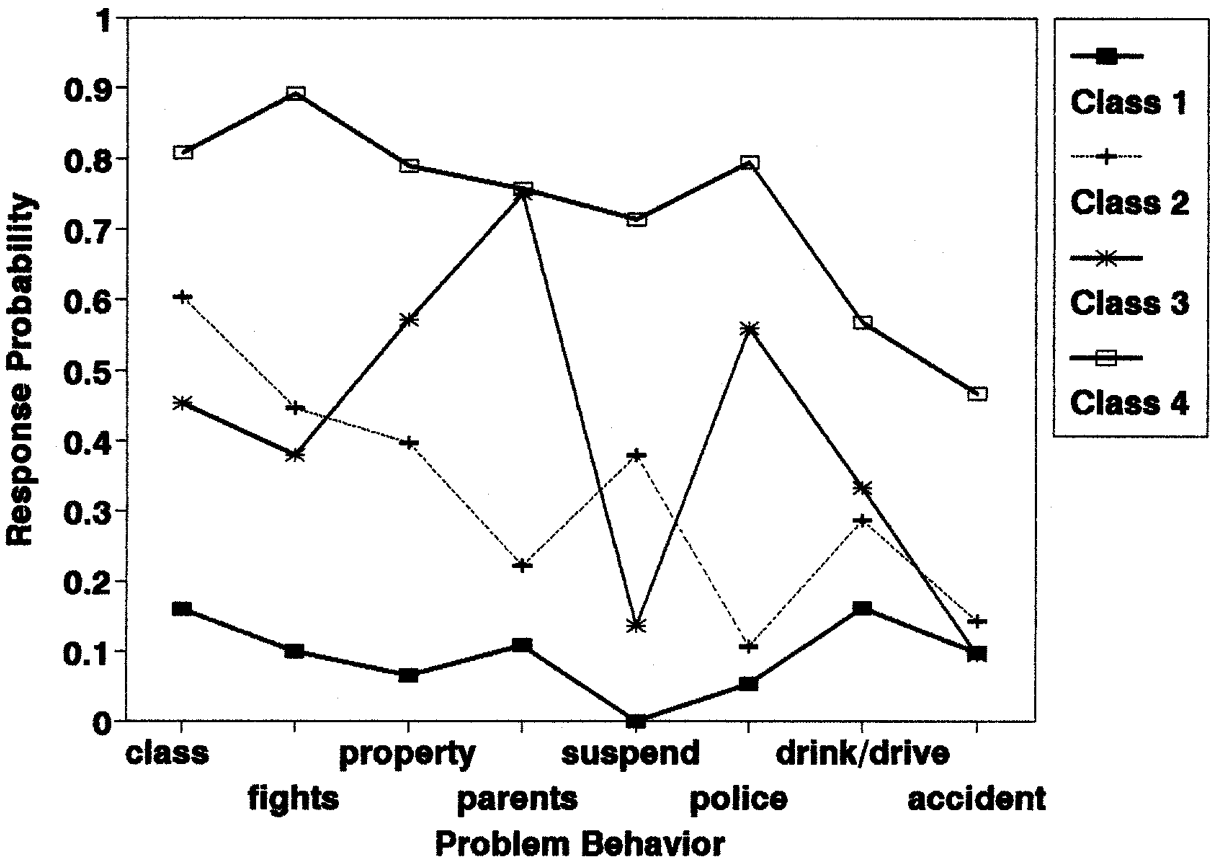

Figure 1: Conditional probabilities (response probabilities) by latent class for four-class model C4.

Insight into the four classes is gained by comparing their patterns of substance use. Substance use is generally high among Class 4 students and low among Class 1 students. The middle classes are approximately equal in terms of wine, cigarette, and marijuana use. However, Class 3 students use more beer, liquor, and – most markedly – smokeless tobacco than their Class 2 counterparts. The picture that emerges with respect to the two middle classes is thus one of trouble with parents and police and property damage, combined with beer, liquor, and smokeless tobacco use, on the one hand, and being sent from the classroom and suspended on the other. Potentially, these represent distinct „syndromes“ of problem behavior, which might require different interventions.

8. Concluding Remarks

The results suggest it may be useful to think in terms of a latent dimension of problem behavior. Further research should aim to see if this construct is associated with other personality, behavioral, educational, and developmental variables. At the same time, it appears useful to distinguish separate groups of students in terms of problem behaviors. Again, further research may establish connections between these classes and other factors.

We have illustrated how it can be useful to apply a variety of latent structure models to the same data. As an exploratory strategy, it is helpful to view the data in alternative ways. We have further demonstrated how certain statistical inferences about latent structure may be obtained by comparing results of different models.

Acknowledgements

This research was supported in part by National Institute on Drug Abuse grant 1-R01-DA07030 and an INVEST grant. The assistance of the Forsysth Country, North Carolina School District is gratefully acknowledged.

References

Bock, R.D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*, 443-459.

Dayton, C.M., & Macready, G.B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association, 83*, 173-178.

Donovan, J.E., & Jessor, R. (1985). Structure of problem behavior in adolescence and young adulthood. *Journal of Consulting and Clinical Psychology, 53*, 890-904.

Everitt, B.S., & Merette, C. (1990). The clustering of mixed-mode data: A comparison of possible approaches. *Journal of Applied Statistics, 17*, 283-297.

Formann, A.K. (1985). Constrained latent class models: Theory and applications. *British Journal of Mathematical and Statistical Psychology, 38*, 87-111.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association, 87*, 476-486.

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

Gottfredson, M.R., & Hirschi, T. (1990). *A general theory of crime*. Stanford: Stanford University Press.

Henkelman, R.M., Kay, I., & Bronskill, M.J. (1990). Receiver operator characteristic (ROC) analysis without truth. *Medical Decision Making, 10*, 24-29.

Jessor, R., & Jessor, S.L. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth*. New York: Academic Press.

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

Lindsay, B., Clogg, C.C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association, 86*, 96-107.

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

Mislevy, R.J. (1984). Estimating latent distributions. *Psychometrika, 49,* 359-381.

Pol, F. van de, Langeheine, R., & de Jong, W. (1989). *PANMARK user manual*. Voorburg, The Netherlands: Netherlands Central Bureau of Statistics.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests, 2nd edition*. Chicago: University of Chicago Press.

Rost, J. (1988). Rating scale analysis with latent class models. *Psychometrika, 53,* 327-348.

Qu, Y., Tan, M. & Kutner, M.H. (1996). Random effects model in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics, 52*, 797-810.

Sclove, S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333-343.

Uebersax, J.S. (1993). Statistical modeling of expert ratings on medical treatment appropriateness. *Journal of the American Statistical Association, 88,* 421-427.

Uebersax, J.S. (1994a). A unifying framework for latent mixture models with dichotomous or ordered-category data. Paper presented at the annual meeting of the Biometric Society, Eastern North American Region, Cleveland, Ohio.

Uebersax, J.S. (1994b). LLCA user's manual. Unpublished computer program documentation.

Uebersax, J.S. (1996). Probit latent class models for dichotomous data. Submitted manuscript.

Uebersax, J.S., & Grove, W.M. (1993). A latent trait finite mixture model for the analysis of rating agreement. *Biometrics, 49*, 823-835.