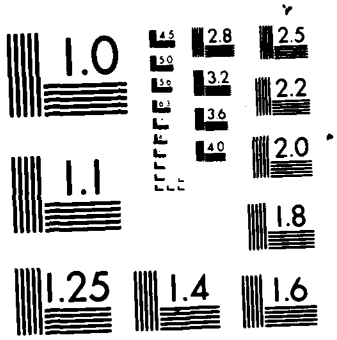
AD-A178 209 A DISCRETE LATENT STATE APPROACH TO DIAGNOSTIC TESTING 1/1
(U) PORTLAND STATE UNIV OR DEPT OF PSYCHOLOGY
J A PAULSON AUG 86 TR-86-2 N00014-81-K-0564

UNCLASSIFIED F/G 5/10 NL

END
DATE
FILMED
4 87



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

(1)

A DISCRETE LATENT STATE APPROACH

TO DIAGNOSTIC TESTING: FINAL REPORT

ON CONTRACT NUMBER N00014-81-K-0564

James A. Paulson

Psychology Department

Portland State University

August 1986

This research was sponsored by the Personnel and Training and Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-81-K-0564, NR 150-466.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	Technical Report ONR 86-2
2. GOVT ACCESSION NO.	AD-A1[illegible]
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle)	A Discrete Latent State Approach to Diagnostic Testing: Final Report on Contract Number N00014-81-K-0564
5. TYPE OF REPORT & PERIOD COVERED	Final Report
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s)	James A Paulson
8. CONTRACT OR GRANT NUMBER(s)	N00014-81-K-0564
9. PERFORMING ORGANIZATION NAME AND ADDRESS	Psychology Department Portland State University Portland, OR 97207
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	NR 150-466
11. CONTROLLING OFFICE NAME AND ADDRESS	Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217
12. REPORT DATE	August 1986
13. NUMBER OF PAGES	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	
15. SECURITY CLASS. (of this report)	Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Latent class models
Marginal maximum likelihood
EM algorithm
Signed-number arithmetic
Systematic response patterns
Monotone homogeneity
Parameter invariance
Componential analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This project has developed the general latent class model as a framework for representation of item responses. This framework can be used to represent data in applications such as mastery tests and other kinds of achievement tests, where there is reason to believe that current foundations are deficient. Methods of estimation for the latent class model have been improved and hypothesis tests addressing issues important in developing specific models for test data have been devised.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

These hypothesis tests include a test for monotone homogeneity of items, tests of invariance of item parameters between groups and over time, a test for the significance of inclusion of a new state in a model, and other tests. A nonparametric approach to maximum likelihood estimation of item response functions for monotonely homogeneous sets of items has been devised. It is easy to generate these tests in principle, because of the ease of dealing with various specifications of fixed, equality, complementarity, and monotone homogeneity constraints in the estimation algorithm.

The use of this general approach has been illustrated by developing models which successfully represent signed-number addition test data gathered by Tatsuoka and Birenbaum (1979). These models are noteworthy because Tatsuoka and Birenbaum have shown (an our new monotone homogeneity test has confirmed) that this data cannot in principle be represented by a unidimensional model. A number of technical issues relating to these models are discussed.

Introduction

The overriding goal of this project has been to develop a general framework for representation of item responses which can be used to represent data in applications such as mastery tests and other kinds of achievement tests, where there is reason to believe that current foundations are deficient. The strategy which I originally proposed for pursuing this goal involved building a model for signed-number addition test data gathered by Tatsuoka and Birenbaum (1979). They have shown that this data cannot be represented by a unidimensional model because of a number of systematic error patterns exhibited by different subgroups of students. The immediate subgoals of the project have been to:

1. validate a finite latent state model which I developed to account for this data;
2. extend this model to deal with change over time; and
3. develop optimal procedures based on the model for testing mastery of the signed-number addition concept.

The first of these subgoals has taken more time to reach than anticipated, but pursuit of it has yielded results of more general applicability than I had originally hoped to obtain. These results, which I described at the October 1984 ONR Contractors' Meeting at ETS, will be discussed further in the next section of the report. The section after that will treat extensions of these results to models which impose monotone homogeneity constraints on the item parameters. The extensions are important because they serve to explicity relate the general latent class model representation to standard item response theory representations of test data and provide a basis for deciding whether or not the latter representations are approriate for a given set of data. The final section of the report will describe some preliminary results concerning extensions of the model to deal with change over time, simultaneous modelling of more than one response component, and some of the implications of these results for testing procedures based on the model.

Latent Class Models for Item Responses

In order to validate the finite latent state models which I had developed for the signed-number addition data, it occurred to me that it would be nice to formulate a more general model which would include my models as special cases. Then, if reasonable estimation procedures and goodness-of-fit indices could be devised for the general model, it would be possible to answer a number of questions about the validity of specific models. It occurred to me that Lazarsfeld's Latent Class Structure models would include my models as special cases. However, there were problems with estimating parameters in latent class models which seemed to limit the applicability of methods associated with them to my problems. The complexity of existing approaches grows exponentially with the number of items. Ten items would be considered a lot; I was dealing with twenty-item tests.

Since many of the interesting implications of my model concern the structure of interitem correlations, I decided to try to estimate parameters by fitting covariance matrices, in the spirit of Jöreskog's analysis of covariance structures. I developed the necessary theoretical formulas and some computer programs to implement this generalized least squares approach and presented them at the October 1983 ONR Contractors' Meeting at the University of Illinois. I noted various difficulties in getting the algorithms to converge and outlined a quasi-Newton algorithm which I hoped would circumvent them.

The proposed approach to parameter estimation was greeted with some skepticism at the Contractors Meeting. It was suggested that I re-examine the literature on maximum likelihood estimation in latent class models. I was not eager to do this, for reasons alluded to above, but it did seem that it might be worth pursuing the EM approach being used by Tsutakawa and Bock on other problems. I did this, and in effect, wound up reinventing Goodman's (1974) algorithm for constrained marginal maximum likelihood estimation in latent class models, but with an essential modification which dramatically extends the algorithm to apply to tests with many items.

The EM approach yields a particularly simple algorithm in the case of the latent class model. The computations on each iteration are straightforward because of the finite number of states. In the expectation phase, or E-phase, of each iteration the conditional state probabilities, given the trial parameter values and the subject's responses, are apportioned to each state according to these state probabilities. Then, in the maximization, or M-phase, the parameter values are revised by computing estimated "sample" proportions of subjects in each state and estimated "sample" proportions passing each item, given the state, based on the results of the E-phase.

One of the difficulties with my earlier approach to estimation was a tendency for the estimates to drift outside the unit interval to which they are constrained by the fact that they are all probabilities. These constraints are always automatically satisfied by the present algorithm. Not only are these constraints satisfied, it is easy to modify the algorithm to require subsets of the item parameters to be equal or complementary to each other. These additional constraints are also automatically satisfied by the nature of the algorithm.

When the maximum likelihood estimates have been obtained, it is easy to compute the marginal likelihood of the data as a whole. By computing likelihoods under hypotheses imposing different constraints, one can perform likelihood ratio tests to answer a variety of questions. When these tests are applied to the signed-number addition data, the specific models which I have proposed are seen to give a qualitatively good account of the data, but they are wrong on some details. For example, the models imply that items within types should be equivalent in the sense of having identical parameters. This equivalence hypothesis must be rejected. The models imply, that in states corresponding to systematic response patterns, the probabilities of deviant responses are the same for all item types. This hypothesis must also be rejected.

While the null hypotheses must be rejected, examination of the unconstrained parameter estimates reveals that the deviations from these hypotheses are relatively minor. If only small samples are available for estimating parameters, as is the case here, the simpler constrained models probably provide a more robust representation of the data than the more general models.

It would have been surprising if these analyses had turned out any differently than they did, because Yamamoto (1983) got very similar results with the same data but different methods. Besides confirming Yamamoto's results, the point of these analyses is that they demonstrate the use of a much more flexible approach to model development questions for latent class models.

During December 9-21, 1984 I participated in the NATO Advanced Study Institute on Human Assessment: Advances in Measuring Cognition and Motivation, in Athens, Greece. I presented a paper entitled "Latent Class Representation of Systematic Patterns in Test Responses," which was basically an account of the work which I have just described above. Since then I have expanded the paper into a general discussion of latent class structure as a framework for modeling test performance, using signed-number addition models to illustrate the process of model development. The paper Paulson (1985), will be published in Irvine, S.H., Newstead, S. and Dann, P. (eds.) Computer-Based Human Assessment, a volume of selected papers from the ASI, to be published by Nijhoff. It is also being distributed as a technical report simultaneously with this Final Report. Five questions which will frequently arise in building latent class models are treated at some length in the paper.

The questions are:

1. How many states should the model have?
2. Are nominally equivalent items really equivalent?
3. Does a given specific parametric model hold?
4. Are the item parameters of a given model invariant over time?
5. Are the item parameters invariant across groups?

Likelihood ratio tests for dealing with each question are described in detail. It is easy to generate these tests in principle, because of the ease of dealing with various specifications of fixed, equality, and complementarity constraints in the estimation algorithm.

Monotone Homogeneity of Items

The likelihood ratio principle has been used to construct a wide variety of hypothesis tests relevant to the development of latent class models. However, one issue which does not lend itself directly to such a test is the basic question of whether a unidimensional latent trait model might adequately account for a given data set. The problem is that neither model is nested in the other: the unidimensional model has an infinite set of states, whereas a finite state latent class model need not be unidimensional. One way out of the problem would be to estimate ICC's for some unidimensional model, such as the three-parameter logistic model, discretize θ at a finite number of points sufficient to represent the curves, use the resulting $\hat{P}_j(\theta_k)$'s as P_{kj}'s for a constrained latent class model, and test to see if a more general latent class model accounts significantly better for the data than the discretized unidimensional model. While this approach may be a good way to test the fit of the particular model chosen, it is not an adequate test of unidimensional models in general. Some other unidimensional model might fit fine, if the model chosen does not. A better approach is suggested by the following observation.

Suppose that we estimate conditional probabilities of correct response to items, given state, in an unconstrained latent class model, and find that the ordering of the P_{kj}'s is the same for all items. That is, the items are "monotonely homogeneous" in the term used by Charles Lewis (1985). If we do, it strongly suggests that an adequate unidimensional model could be found. However, if we find instead that the deviations from monotonicity can not be attributed to sampling variability, it implies that no such unidimensional model can be found.

Nonparametric estimation of monotonely homogeneous ICC's. A simple extension of the algorithms developed to deal with equality constraints can provide marginal maximum likelihood estimates of the parameters in a latent class model, subject only to the constraint of monotone homogeneity of the item parameters. The fact that the monotonely constrained model is nested in the unconstrained model with the same number of states leads directly to a likelihood ratio test of monotone homogeneity. If the monotone homogeneity hypothesis is acceptable, the constrained parameter estimates for each state plotted against expected number of items correct, given state, provides nonparametric marginal maximum likelihood estimates of the ICC's.

Due to the finite number of states, this approach can only yield an approximation to the ICC's. However, if one uses enough states this should not be too much of a problem. I think it would be very interesting to compare this approach to other approaches which make no assumptions regarding the form of the ICC. The approach is promising because there is a very simple way to accommodate the monotone homeogeneity constraint.

The "Up-and-Down Blocks" algorithm. Consider a simpler problem than the present one. We have responses to a given item from individuals in a series of groups, and we assume the groups fall in a known order with respect to probability of correct response to the item. What is the maximum likelihood estimator of the set of group probabilities, subject to the ordering constraint? Without the constraint, the MLE is just the set of sample proportions correct in each group. If the sample proportions happen to fall in the assumed order, the constraint is not active and the unconstrained MLE applies. If the sample proportions do not all fall in the prescribed order, then a theorem from the theory of isotonic regression says how the constrained MLE can be constructed from the sample proportions by amalgamating groups into level sets within which equality constraints apply. The "Up-and-Down Blocks" algorithm is a simple procedure devised by Kruskal (1964) for effecting this division into level sets. These developments are described in detail by Barlow, Bartholomew, Bremner, and Brunk (1972). Since my program can handle equality constraints, and the P_{jk}'s yielded by the unconstrained phase of each iteration of my program are analogous to sample proportions correct in the respective states, the extension to the monotonely homogeneous constraints is straightforward.

The test of monotone homogeneity. If there are J items on a test and one is fitting an unconstrained latent class model with s states, then there are Js free item parameters to be estimated. Let m_j denote the number of level sets determined by the Up-and-Down Blocks algorithm for item j. The number of free item parameters in the model with the monotone homogeneity constraint is then $\sum m_j$. Let L_u and L_m denote the maxima of the likelihood function evaluated under the unconstrained and monotonely constrained hypothesis, respectively. If the monotone homogeneity hypothesis is correct, then asymptotically the likelihood ratio test statistic

$$-2 \log \lambda = 2(\log L_u - \log L_m)$$

has a chi squared distribution with $J\underline{s} - \sum_j m$ degrees of freedom.

This fact can be used to set up critical regions for tests of the hypothesis. A detailed discussion of the extension of the EM approach to deal with monotone homogeneity constraints is given in Paulson (1986), a technical report which is being distributed simultaneously with this Final Report.

Some Important Technical Questions

This section describes the results of some preliminary analyses which might help answer the following questions regarding signed-number addition test performance:

1. Are items parameters invariant from one testing to the next?
2. Are the states into which subjects are classified on different response components related? If so, can a simple model be found relating the distribution on the joint classification to the marginal distributions on the separate components?
3. How do subjects move from state to state during the course of learning?

If the item parameters are invariant over time, then changes in performance can be interpreted as transitions between states; if they are not, the interpretation of change is problematical. Even if the changes in parameters over time are relatively minor deviations which do not affect the qualitative interpretations of the states, parameter dependent statistical procedures for characterizing test performance might be adversely affected by them.

In a completely satisfactory componential model for test responses, the number of states needed to characterize the responses is the product of the numbers of states in the models for the respective components. Accurate estimation of parameters in the comprehensive model is not likely to be feasible unless a simple model relating the joint distribution over states to the marginal distributions can be found.

The question regarding the transitions between states which subjects make during the course of learning makes sense even if the nature of the states changes from points early in learning to points late in learning. Which transitions occur most often might well have pedagogical significance. It may also have theoretical implications for methods of assessing change.

The data to be presented in addressing these questions comes from a panel of Junior High School students in Urbana, Illinois who were studied by Tatsuoka and Birenbaum. Most of them took their first signed-number arithmetic test at the same time as the students discussed by Tatsuoka and Birenbaum (1979), whose data I analyzed in detail in Paulson (1985). When first tested, the students had only received a small amount of experimental instruction on signed-numbers. As was expected, many of them still did not understand signed-number addition after this brief exposure. The panel of students was next tested at the beginning of regular classroom instruction on signed-numbers, after an interval of some weeks. Thus, the second test was essentially a retention test. There is data on 59 students at this second testing. Two of our analyses involve data on the second test; a third analysis involves the relationship between performance on the first and second tests. There is also data available for many of these subjects from two tests later in instruction. This data will not be presented here in detail, because the number of subjects who did not master signed-number addition before the later tests was too small.

Parameter invariance. Table 1 gives parameter estimates based on data on the magnitude response component from subjects on the second test for two different models. Both models assume that all items of a given type have identical parameter values. The first model constrains the item parameters to be equal to the estimates based on the data from the first testing. Only the parameters giving the distribution of subjects over states are reeestimated using the data from the second test. The second model reestimates all the parameters. The likelihood-ratio chi-squared statistic for testing the hypothesis that the second set of item parameters is identical to the first set is highly significant: $\chi^2(25)=69.46$, $p<.001$. Hence, the hypothesis of parameter invariance must be rejected. Examination of Table 1 reveals, however, that none of the differences between the item parameters obtained on the two occasions affects *the qualitative interpretations* of the patterns of responses to different item types in the various states. All of the differences are less than .20 and the only parameter values which change from less than .50 to greater than .50, or vice versa, are those which fall in the .40-.60 range for both models.

Table 1 about here

Some of the differences which contribute to the significant chi-squared statistic are the following. In the model with item parameters constrained to equal their values on the first test, relatively more of the subjects would be classified as belonging to the random response state and relatively fewer to the systematic response states than would be so classified in the model with recalibrated item

Table 1. Comparison of item parameter estimates for the magnitude response component, based on tests of the same subjects on two different occasions (N=59).

State		State Probability		Item Type									
				L+-S		-S+L		-L+-S		S+-L		-L+S	
		t1*	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
Mastery	1.	.21	.24	.88	.91	.88	.70	.75	.83	.92	.84	.94	.95
Systematic Error Patterns (2 - 4)	2.	.16	.17	.94	.95	.96	.98	.05	.07	.79	.97	.85	.88
	3.	.09	.15	.15	.13	.20	.04	.80	.88	.02	.14	.00	.09
	4.	.02	.02	.82	.75	.25	.25	.20	.00	.95	1.00	.12	.50
Random	5.	.52	.42	.55	.41	.48	.52	.46	.27	.32	.29	.49	.55

Test of goodness-of-fit of model employing time 1 item parameter values on data obtained at time 2: $\chi^2(25)=69.46$, $p<.001$.

*NOTE: This is the estimated state distribution at time 2, using the constrained model with time 1 item parameters, not the state distribution at time 1.

parameters. When the item parameters are reestimated, a few more subjects appear to have mastered the component and a few more appear to fall into the "almost always add" error pattern. Other things being equal, response patterns in the random state tend to have smaller marginal likelihoods than patterns typical of systematic states represented in the model. Levine and Drasgow (1980) make a similar observation in connection with appropriateness measurement: in latent trait models, the conditional likelihoods of response patterns, given the maximum likelihood estimate of Θ for the response pattern, tend to increase with Θ. This explains how the moderate deviations between the item parameters on the two occasions lead to the substantial goodness-of-fit statistic which we get.

The relationship between components. The rest of the analyses we shall report are based on the frequency distributions over combinations of states in component-by-component crossclassification tables. Since the states are not directly observable, we have to resort to indirect means to obtain these frequency tabulations. Rather than adding one tally to the appropriate cell for each individual, we apportion the one tally for each subject to cells on the basis of the likelihoods of respective states, given the response pattern for each individual. To simplify matters, we assume that, conditional upon the response pattern, the states on the respective components are independent. Hence, we add the product of the likelihoods of the states on the components comprising each cell to each of the cells in the table. This yields a table of expected frequencies, which are usually fractions. The estimated joint distribution of subjects over states on the sign and magnitude components on the second test is given in Table 2.

Insert Table 2 about here

The phi-coefficients between the mastery/non-mastery dichotomies on the sign and magnitude components are significantly greater than zero on all tests, falling in range .48 to .62. Thus, the simplest model for the joint distribution, which assumes that classifications on the two components are independent, fails on every testing. A simple model which takes the association between mastery on the two components into account, but implies conditional independence, *given that one* component or the other has not been mastered, can be specified as follows. Let π_{ij} be the joint probability of a subject being in state i on the sign component and state j on the magnitude component. Let $\pi_{i.}$ and $\pi_{.j}$ denote the corresponding marginal probabilities of states on the components, and let λ denote the covariance between the two mastery/non-mastery dichotomies. Then the simple dependence model is given by

$$\pi_{ij} = \begin{cases} \pi_{1.}\pi_{.1}\left(1+\dfrac{\lambda}{\pi_{1.}\pi_{.1}}\right) & \text{for } i=1,\ j=1; \\[2ex] \pi_{1.}\pi_{.j}\left(1-\dfrac{\lambda}{\pi_{1.}(1-\pi_{.1})}\right) & \text{for } i=1,\ j>1; \\[2ex] \pi_{j.}\pi_{.1}\left(1-\dfrac{\lambda}{\pi_{.1}(1-\pi_{1.})}\right) & \text{for } i>1,\ j=1; \\[2ex] \pi_{i.}\pi_{.j}\left(1+\dfrac{\lambda}{(1-\pi_{1.})(1-\pi_{.1})}\right) & \text{for } i>1,\ j>1. \end{cases}$$

Table 2. Estimated joint distribution of subjects over states on the <u>sign</u> and <u>magnitude</u> components at the second testing. Frequencies <u>expected</u> under a <u>simple</u> *model* for dependence are given in parentheses below the empirical frequencies.

State on Sign Component		Mastery 1	Systematic Error Patterns (2 - 4) 2	3	4	Random 5	Row Total
			State or Magnitude of Component				
Mastery	1.	7.8 (7.8)	.8 (.4)	.0 (.3)	.0 (.3)	.9 (.9)	9.5
Systematic Error Patterns (2 - 5)	2.	.0 (.7)	2.7 (1.1)	2.0 (1.1)	.0 (.1)	1.2 (2.9)	5.9
	3.	1.4 (1.4)	2.0 (2.1)	3.7 (2.0)	.0 (.2)	4.1 (5.4)	11.2
	4.	.0 (.5)	.0 (.8)	1.0 (.7)	.0 (.1)	3.0 (1.9)	4.0
	5.	.0 (.1)	.0 (.2)	.0 (.2)	.0 (.0)	1.0 (.5)	1.0
Random	6.	4.8 (3.4)	4.4 (5.3)	2.3 (4.8)	1.0 (.5)	14.9 (13.3)	27.4
COLUMN TOTAL		14.0	9.9	9.0	1.0	25.1	59.0

Test of goodness-of-fit of simple dependence model: $\chi^2(19)=13.61$, $p>.75$.

In this notation, state 1 is the mastery state on both components.

Yamamoto (1983) showed that this simple dependence model fits the data from the first test quite well. In fact, a restricted form of the model in which mastery of the magnitude component implies mastery of the sign component, gives a satisfactory account of the data.

The data in Table 2 show that the restricted form of the model for dependence breaks down on the retention test (Test 2), because several subjects who appear to have mastered the magnitude component have not mastered the sign component. Nevertheless, the general form of the model fits the data very well.

The simple dependence model does a pretty good job of accounting for the data on the third and fourth tests also. Only on the fourth test does it show any sign of breaking down. Two subjects on that test form a class by themselves on both components: on the sign component, they tend to take the sign of the second addend as the sign of the sum; on the magnitude component, they tend to subtract when the sign of the second addend is negative and add otherwise. As was indicated above, most subjects have mastered both components by the fourth test. Only 11 cells in the 5x6 contingency table have expected frequencies greater then 1 under the simple dependence model, so the appropriateness of the goodness-of-fit test is subject to question. The distribution of the other 76 of the 78 subjects who took the test was in good accord with the simple dependence model. Under most circumstances a significant test statistic which is entirely due to 2 observations falling in a cell with very small expected frequency should be viewed with skepticism. Certainly, the model represents most of the data well. In this case, however, the "outliers" make good psychological sense and serve to demonstrate how the model would be likely to break down in practice. It would be a mistake to ignore them.

Transitions between states. Examination of the joint distribution of subjects' states on the magnitude component on the first two tests, given in Table 3, shows that approximately three-fourths of the subjects either stayed in the state they were in on the first test or moved to the random state. This pattern applies to transitions from the mastery state on the first test and to the transitions from all but one of the systematic error states as well. The same tendency appears in the transitions from state to state between the second and third test and between the third and fourth test, except that transitions to the mastery state become the most common transition from every state after classroom instruction begins.

Insert Table 3 about here

These results have a certain verismilitude in the context of the latent class model. The fact that many subjects have similar response patterns on both tests lends credibility to our qualitative interpretations of these response patterns as states. The fact that many other subjects shift from systematic responding to more or less random responding after a period of no instruction on signed-numbers would probably not come as a surprise to their teachers. While these results make sense in terms of the latent class model, it might be

Table 3. Joint distribution of subjects' states on the magnitude response component on the first and second tests. Estimates of the transition probabilities from states on the first test to states on the second test are given in parentheses below the frequencies in the cross-tabulation.

State on First Test		State on the Second Test					
		Mastery	Systematic Error Patterns (2 - 4)			Random	Row Total
		1	2	3	4	5	
Mastery	1.	4.3 (.40)	.4 (.04)	.3 (.02)	.7 (.06)	5.1 (.47)	10.8
Systematic Error Patterns (2 - 4)	2.	1.0 (.07)	6.5 (.44)	.8 (.06)	.3 (.02)	6.2 (.42)	14.8
	3.	1.0 (.17)	.0 (.00)	2.3 (.39)	.0 (.00)	2.6 (.44)	5.9
	4.	2.9 (.46)	1.0 (.16)	1.0 (.16)	.0 (.00)	1.4 (.22)	6.4
Random	5.	1.9 (.14)	.0 (.00)	2.7 (.21)	.0 (.00)	8.5 (.65)	13.1
COLUMN TOTAL		11.1	7.9	7.1	1.0	23.9	51.0

noted that they would not be well represented by the statistical models usually employed in assessing change. The latter models are implicitly or explicitly unidimensional. If individual differences in the amount of change are allowed for at all, they are regarded as random effects. At least in the present instance, latent class models provide a richer, and apparently more valid, representation of the changes.

Summary

This project has developed the general latent class model as a framework for representation of item responses. This framework can be used to represent data in applications such as mastery tests and other kinds of achievement tests, where there is reason to believe that current foundations are deficient. Methods of estimation for the latent class model have been improved and hypothesis tests addressing issues important in developing specific models for test data have been devised.

These hypothesis tests include a test for monotone homogeneity of items, tests of invariance of item parameters between groups and over time, a test for the significance of inclusion of a new state in a model, and other tests. A nonparametric approach to maximum likelihood estimation of item response functions for monotonely homogeneous sets of items has been devised. It is easy to generate these tests in principle, because of the ease of dealing with various specifications of fixed, equality, complementarity, and monotone homogeneity constraints in the estimation algorithm.

The use of this general approach has been illustrated by developing models which successfully represent signed-number addition test data gathered by Tatsuoka and Birenbaum (1979). These models are noteworthy because Tatsuoka and Birenbaum have shown (and our new monotone homogeneity test has confirmed) that this data cannot in principle be represented by a unidimensional model. A number of techincal issues relating to these models are discussed.

References

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. Statistical inference under order restrictions. New York: John Wiley, 1972.

Goodman, L. A. The analysis of qualitative variables when some of the variables are unobservable. Part I--A modified latent structure appraoch. American Journal of Sociology, 1974, 79, 1179-1259.

Kruskal, J. B. Nonmetric multidimensional scaling: a numerical method. Psychometrika, 1964, 29, 115-129.

Levine, M. and Drasgow, F. Appropriateness measurement: basic principles and validating studies. In Weiss, D. J. (ed.) Proceedings of the 1979 Computerized Adaptive Testing Conference, Department of Psychology, University of Minnesota, Minneapolis, MN, September 1980.

Lewis, C. Developments in nonparametric ability estimation. In Weiss, D. J. (ed.) Proceedings of the 1982 Item Response Theory and Computerized Adaptive Testing Conference, Department of Psychology, University of Minnesota, Minneapolis, MN, April 1985.

Paulson, J. A. Latent class representation of systematic patterns in test responses. Technical Report ONR85-1, Portland State University, Department of Psychology, May 1985.

Paulson, J. A. Estimation of parameters in latent class models with constraints on the parameters. Technical Report ONR 86-1, Portland State University, Department of Psychology, Portland, OR, June 1986.

Tatsuoka, K., and Birenbaum, M. The danger of relying soley on diagnostic adaptive testing when prior and subsequent instructional methods are different. Computer-based Education Research Laboratory Report E-5. University of Illinois, 1979.

Yamamoto, K. Derivation and test of predictions of a discrete latent state model for signed number addition test performance. Master's thesis, Department of Psychology, Portland State University, Portland, OR, 1983.

Personnel Analysis Division,
AF/MPXA
5C360, The Pentagon
Washington, DC 20330

Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Phipps Arabie
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Technical Director, ARI
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N711
Naval Training Equipment Center
Orlando, FL 32813

Dr. R. Darrell Bock
University of Chicago
Department of Education
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code N-095R
NAVTRAEQUIPCEN
Orlando, FL 32813

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. Patricia A. Butler
NIE Mail Stop 1806
1200 19th St., NW
Washington, DC 20208

Mr. James W. Carey
Commandant (G-PTE)
U.S. Coast Guard
2100 Second Street, S.W.
Washington, DC 20593

Dr. James Carlson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

Dr. Robert Carroll
NAVOP 01B7
Washington, DC 20370

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

1986/03/19

Portland State University/Paulson

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans Crombag
University of Leyden
Education Research Center
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

CTB/McGraw-Hill Library
2500 Garden Road
Monterey, CA 93940

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Ball Foundation
800 Roosevelt Road
Building C, Suite 206
Glen Ellyn, IL 60137

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. John M. Eddins
University of Illinois
252 Engineering Research
Laboratory
103 South Mathews Street
Urbana, IL 61801

Dr. Susan Embretson
University of Kansas
Psychology Department
Lawrence, KS 66045

ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52240

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

Portland State University/Paulson

Dr. Carl H. Frederiksen
McGill University
3700 McTavish Street
Montreal, Quebec H3A 1Y2
CANADA

Dr. Robert D. Gibbons
University of Illinois-Chicago
P.O. Box 6998
Chicago, IL 69680

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Dr. Ronald K. Hambleton
Prof. of Education & Psychology
University of Massachusetts
at Amherst
Hills House
Amherst, MA 01003

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 62
San Diego, CA 92152

Dr. Paul W. Holland
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Universitat Dusseldorf
Erziehungswissenschaftliches
Universitatsstr. 1
Dusseldorf 1
WEST GERMANY

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 90010

Mr. Dick Hoshaw
NAVOP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. Douglas H. Jones
Advanced Statistical
Technologies Corporation
10 Trafalgar Court
Lawrenceville, NJ 08148

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
University of Texas-Austin
Measurement and Evaluation
Center
Austin, TX 78703

Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
The NETHERLANDS

Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

Dr. William L. Maloy
Chief of Naval Education
and Training
Naval Air Station
Pensacola, FL 32508

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Dr. James McBride
Psychological Corporation
c/o Harcourt, Brace,
Javanovich Inc.
1250 West 6th Street
San Diego, CA 92101

Dr. Clarence McCormick
HQ, MEPCOM
MEPCT-P
2500 Green Bay Road
North Chicago, IL 60064

Mr. Robert McKinley
University of Toledo
Department of Educational Psychology
Toledo, OH 43606

Dr. Barbara Means
Human Resources
Research Organization
1100 South Washington
Alexandria, VA 22314

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Headquarters, Marine Corps
Code MPI-20
Washington, DC 20380

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

Dr. William E. Nordbrock
FMC-ADCO Box 25
APO, NY 09710

Dr. Melvin R. Novick
356 Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152

Library, NPRDC
Code P201L
San Diego, CA 92152

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

Office of Naval Research,
Code 1142PT
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Special Assistant for Marine
Corps Matters,
ONR Code 00MC
800 N. Quincy St.
Arlington, VA 22217-5000

Dr. Judith Orasanu
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dr. Roger Pennell
Air Force Human Resources
Laboratory
Lowry AFB, CO 80230

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
Knoxville, TN 37916

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152

Dr. Robert Sasmor
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujusawa 251
JAPAN

Dr. William Sims
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Paul Speckman
University of Missouri
Department of Statistics
Columbia, MO 65201

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801

Maj. Bill Strickland
AF/MPXOA
4E168 Pentagon
Washington, DC 20330

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
Urbana, IL 61801

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
The Fred Hutchinson
Cancer Research Center
Division of Public Health Sci.
1124 Columbia Street
Seattle, WA 98104

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E. Street, NW
Washington, DC 20415

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 310
St. Paul, MN 55114

Dr. Frank Vicino
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08541

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Mr. Thomas A. Warm
Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Dr. Brian Waters
Program Manager
Manpower Analysis Program
HumRRO
1100 S. Washington St.
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

Dr. Ronald A. Weitzman
NPS, Code 54Wz
Monterey, CA 92152

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90007

German Military Representative
ATTN: Wolfgang Wildegrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

ATE

LMED

-8

[illegible]
ISRAEL

Dr. Arthur S. Blaiwes
Code N711
Naval Training Equipment Center
Orlando, FL 32813

Dr. Robert Carroll
NAVOP 01B7
Washington, DC 20370

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007