DOCUMENT RESUME

ED 384 678 TM 024 000

AUTHOR	Mislevy, Robert J.; Wilson, Mark
TITLE	Marginal Maximum Likelihood Estimation for a Psychometric Model of Discontinuous Development.
INSTITUTION	California Univ., Berkeley. Graduate School of Education.; Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Office of Naval Research, Arlington, VA. Cognitive and Neural Sciences Div.; Spencer Foundation, Chicago, Ill.
REPORT NO	ETS-RR-92-74-ONR
PUB DATE	Dec 92
CONTRACT	N00014-88-K-0304; PE-61153N; PR-RR-04204; TA-RR-04204-01; WU-R&T-4421552
NOTE	50p.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Bayesian Statistics; *Change; *Development; *Item Response Theory; Learning; *Maximum Likelihood Statistics; Probability; Psychological Characteristics; *Psychometrics; Simulation; Test Items
IDENTIFIERS	EM Algorithm; *Marginal Maximum Likelihood Statistics; *Saltus Model

ABSTRACT

Standard item response theory (IRT) models posit latent variables to account for regularities in students' performance on test items. They can accommodate learning only if the expected changes in performance are smooth, and, in an appropriate metric, uniform over items. Wilson's "Saltus" model extends the ideas of IRT to development that occurs in stages, where expected changes can be discontinuous, show different patterns for different types of items, and even exhibit reversals in probabilities of success on certain tasks. Examples include Piagetian stages of psychological development and Siegler's rule-based learning. This paper derives marginal maximum likelihood (MML) estimation equations for the structural parameters of the Saltus model and suggests a computing approximation based on the EM algorithm. For individual examinees, Empirical Bayes probabilities of learning-stage are given, along with proficiency parameter estimates conditional on stage membership. The MML solution is illustrated with simulated data and an example from the domain of mixed number subtraction. (Contains 29 references, 8 tables, and 1 figure.) (Author)

RR-92-74-ONR

MARGINAL MAXIMUM LIKELIHOOD ESTIMATION FOR A PSYCHOMETRIC MODEL OF DISCONTINUOUS DEVELOPMENT

Robert J. Mislevy
Educational Testing Service

Mark Wilson
University of California, Berkeley

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-88-K-0304
R&T 4421552

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, New Jersey

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December, 1992	Final

4. TITLE AND SUBTITLE
Marginal Maximum Likelihood Estimation for a Psychometric Model of Discontinuous Development

5. FUNDING NUMBERS
G. N00014-88-K-0304
PE 61153N
PR RR 04204
TA RR 04204-01
WU R&T 4421552

6. AUTHOR(S)
Robert J. Mislevy & Mark Wilson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Educational Testing Service // Graduate School of Educ.
Rosedale Road, Princeton, NJ 08541 // University of California, Berkeley, CA

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Sciences
Code 1142CS
Office of Naval Research
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified/Unlimited

12b. DISTRIBUTION CODE
N/A

13. ABSTRACT *(Maximum 200 words)*

Standard item response theory (IRT) models posit latent variables to account for regularities in students' performances in test items. They can accomodate learning only if the expected changes in performance are smooth and, in an appropriate metric, uniform over items. Wilson's "Saltus" model extends the ideas of IRT to development that occurs in stages, where expected changes can be discontinuous, show different patterns for different types of items, and even exhibit reversals in probabilities of success in certain tasks. Examples include Piagetian stages of psychological development and Siegler's rule-based learning. This paper derives mariginal maximum likelihood (MML) estimation equations for the structural parameters of the Saltus model and suggests a computing approximation based on the EM algorithm. For individual examinees, Empirical Bayes probabilities of learning-stage are given, along with proficiency parameter estimates conditional on stage membership. The MML solution is illustrated with simulated data and an example from the domain of mixed number subtraction.

14. SUBJECT TERMS
Cognitive diagnosis, empirical Bayes, item response theory, marginal maximum likelihood, mixture models, Saltus model

15. NUMBER OF PAGES
43 + RDP

16. PRICE CODE
N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C -	Contract	**PR** -	Project
G -	Grant	**TA** -	Task
PE -	Program Element	**WU** -	Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

Marginal Maximum Likelihood Estimation for a Psychometric Model of Discontinuous Development

Robert J. Mislevy

Educational Testing Service

and

Mark Wilson

Graduate School of Education

University of California, Berkeley

The authors' names appear in alphabetical order. We would like to thank Karen Draney for computer programming, Kikumi Tatsuoka for allowing us to use the mixed-number subtraction data, and Kikumi Tatsuoka and Chan Dayton for helpful suggestions. The first author's work was supported by Contract No. N00014-88-K-0304, R&T 4421552, from the Cognitive Sciences Program, Cognitive and Neural Sciences Division, Office of Naval Research, and by the Program Research Planning Council of Educational Testing Service. The second author's work was supported by a National Academy of Education Spencer Fellowship and by a Junior Faculty Research Grant from the Committee on Research, University of California at Berkeley.

Marginal Maximum Likelihood Estimation for a Psychometric Model of Discontinuous Development

Abstract

Standard item response theory (IRT) models posit latent variables to account for regularities in students' performances on test items. They can accommodate learning only if the expected changes in performance are smooth and, in an appropriate metric, uniform over items. Wilson's "Saltus" model extends the ideas of IRT to development that occurs in stages, where expected changes can be discontinuous, show different patterns for different types of items, and even exhibit reversals in probabilities of success on certain tasks. Examples include Piagetian stages of psychological development and Siegler's rule-based learning. This paper derives marginal maximum likelihood (MML) estimation equations for the structural parameters of the Saltus model and suggests a computing approximation based on the EM algorithm. For individual examinees, Empirical Bayes probabilities of learning-stage are given, along with proficiency parameter estimates conditional on stage membership. The MML solution is illustrated with simulated data and an example from the domain of mixed number subtraction.

Key words: Cognitive diagnosis, empirical Bayes, item response theory, marginal maximum likelihood, mixture models, Saltus model

1.0 Introduction

The models of classical test theory and item response theory (IRT) characterize examinees simply in terms of their propensities to make correct answers in a domain of items—that is, their overall proficiencies. Correspondingly, the processes and the outcomes of learning can be expressed through these models only as changes in overall proficiency. This characterization falls short for problems of description and decision-making cast in the framework of what we are learning about how people solve problems, acquire knowledge, and increase their proficiencies (Glaser, 1981; Masters & Mislevy, 1993; Snow & Lohman, 1989). Learners become more competent not simply by accreting additional facts and skills, but by reconfiguring their previous knowledge, by "chunking" information to reduce memory loads, and by developing strategies and models that help them discern when and how facts and skills are relevant. When evaluating or planning instruction, the important questions may not be "How many items did this student answer correctly?" or "What proportion of the population would have scores lower than hers?", but, in Thompson's (1982) words, "What can this person be thinking so that his actions make sense from his perspective?" and "What organization does the student have in mind so that his actions seem, to him, to form a coherent pattern?" Taking this point of view, Glaser, Lesgold, and Lajoie (1987) advocate "achievement testing as ... a method of indexing stages of competence through indicators of the level of development of knowledge, skill, and cognitive process."

Models that incorporate this perspective have begun to appear in the testing literature. Examples include Tatsuoka's (1983, 1990) extension of IRT to "rule space" through the use of cognitive task analyses, Embretson's (1985) and Samejima's (1983) models for alternative response strategies when subtask results can be observed, and Falmagne's (1989), Haertel's (1984), and Paulson's (1986) latent-class models built around the combinations of skills that tasks demand.

Wilson's (1984, 1989) "Saltus" model for learning that occurs in conceptual or developmental stages is another model of this type. Each subject is characterized by two variables, one qualitative and the other quantitative. The qualitative parameter, denoting stage membership, indicates the *nature* of proficiency, while the quantitative parameter indicates *degree* of proficiency. Although both types of parameters are unobservable, approximate solutions in early demonstrations of Saltus treated estimates of stage membership (based on total scores) as if they were known, true, parameter values, followed by "tailored simulations" to correct for some of the effects of this oversimplification. The solution offered in the present paper more properly accounts for the uncertainty associated with examinees' stage memberships, using Mislevy and Verhelst's (1990) empirical Bayesian approach for mixtures of test theory models. After reviewing the form of the Saltus model, we present marginal maximum likelihood (MML) estimation procedures and illustrate their use with simulated data and Tatsuoka's mixed number subtraction data (Klein, Birenbaum, Standiford, and Tatsuoka, 1981).

2.0 The Saltus Model

Wilson's (1984, 1989) Saltus model for hierarchical development generalizes the Rasch model for dichotomous test items (Rasch, 1960/1980) by positing H "developmental stages." An examinee is assumed to be in exactly one stage at the time of testing, but stage membership is not directly observed. Items are also classified into H classes. It is assumed that a Rasch model holds within each developmental stage, and the relative distances between items within a given item class are the same irrespective of developmental stage. The relative difficulties among item classes may differ from one developmental stage to another, however. The amounts by which item class difficulties vary for different stages are the "Saltus parameters." Saltus parameters can capture how certain types of items become much easier relative to others as students reconceptualize a domain or add a new rule to their repertoire, or how certain items can actually become harder as students progress from an earlier stage to a more advanced one if they were previously answered correctly for the wrong reason. Wilson's (1989) illustrative examples concerned the development of children's proportional reasoning abilities, using balance-beam data collected by Siegler (1981), and the acquisition of subtraction rules in a Gagnéan learning hierarchy (see Gagné, 1968).

Anticipating MML estimation, we describe an estimation model in two phases. First is the Saltus *item response model*, which gives probabilities of correct response conditional on stage membership and proficiency. Second is a *population model*, which concerns the proportions of a population of examinees at each stage and the distributions of proficiency within stages.

2.1 The Saltus Item Response Model

Saltus is an extension of the Rasch model (RM) for dichotomous test items. Under the RM, the probability that an examinee with proficiency θ will respond correctly to Item j ($x_j=1$ rather than $x_j=0$) is given as

$$P(x_j=1 \mid \theta, \beta_j) = \Psi(\theta-\beta_j), \tag{1}$$

where β_j is the difficulty parameter of Item j, and Ψ is the cumulative logistic distribution function; that is,

$$\Psi(z) = \exp(z)/[1+\exp(z)]. \tag{2}$$

Under Saltus, an examinee is characterized by not just a proficiency parameter θ, but also a stage membership parameter $\boldsymbol{\phi}$. If there are H potential developmental stages, $\boldsymbol{\phi}_i = (\phi_{i1}, \ldots, \phi_{iH})$, where ϕ_{ih} takes the value of 1 if Examinee i is in Stage h and 0 if not. As with θ, values of $\boldsymbol{\phi}$ are not observable.

Under Saltus, as under the RM, item j has a difficulty parameter β_j. Item j is also associated with developmental stages through the item-class indicator $\mathbf{b}_j$. In analogy to $\boldsymbol{\phi}$,

$\mathbf{b}_j = (b_{j1}, \ldots, b_{jH})$, where b_{jk} takes the value of 1 if item j belongs to item Class k, and 0 otherwise. In contrast with $\boldsymbol{\phi}$, however, $\mathbf{b}_j$ is known a priori for all items.

$\mathbf{T} = (\tau_{hk})$ is an H-by-H matrix of Saltus parameters. In particular, τ_{hk} expresses an effect on the difficulty of items in Class k that applies to examinees in Stage h. The probability that an examinee with stage membership parameter ϕ and proficiency θ will respond correctly to item j is given as

$$P(x_j=1|\theta,\ \boldsymbol{\phi},\ \beta_j,\ \mathbf{T}) = \prod_h \prod_k \Psi(\theta-\beta_j+\tau_{hk})^{\phi_h b_{jk}}. \tag{3}$$

In the sequel, $\Psi(\theta-\beta_j+\tau_{hk})$ will be abbreviated as $\Psi_{jkh}(\theta)$. Note that the double product over h and k in (3) is merely a device to pick up the appropriate Saltus parameter for item j that corresponds to the developmental stage of this particular examinee, since the exponent $\phi_h b_{jk}$ is one in that case and zero otherwise.

Item responses are assumed to be independent given θ and $\boldsymbol{\phi}$. Letting $\mathbf{x} = (x_1, \ldots, x_n)$ be a vector of responses to n items,

$$P(\mathbf{x}\,|\theta,\ \boldsymbol{\phi},\ \boldsymbol{\beta},\ \mathbf{T}) = \prod_j \prod_h \prod_k \{\Psi_{jhk}(\theta)^{x_j}[1-\Psi_{jhk}(\theta)]^{(1-x_j)}\}^{\phi_h b_{jk}}. \tag{4}$$

For brevity, we define

$$P_h(\mathbf{x}\,|\,\theta,\ \beta_j,\ \mathbf{T}) = \prod_j \prod_k \{\Psi_{jhk}(\theta)^{x_j}[1-\Psi_{jhk}(\theta)]^{(1-x_j)}\}^{b_{jk}};$$

$P_h(\mathbf{x}\,|\theta, \boldsymbol{\beta}, \mathrm{T})$, or $P_h(\mathbf{x}\,|\theta)$ for short, is the conditional probability of a response pattern $\mathbf{x}$ given θ *and* membership in Stage h.

2.1.1 Restrictions to Resolve Scaling Indeterminacies

The model defined in (3) is not identified unless further restrictions are imposed on item and Saltus parameters. This can be accomplished in several ways, but once parameters have been estimated under one set of restrictions, it is straightforward to translate them to what they would be under a different set. The following restrictions prove convenient for MML estimation:

$$\Sigma \beta_j = 0,$$

so that item parameters are centered around the origin;

$$\tau_{1k} = 0 \text{ for all k},$$

so that the item parameter estimates apply directly to Stage 1 in a simple RM, but relative changes in item difficulties may apply for other stages via Saltus parameters; and

$$\tau_{h1} = 0 \text{ for all h},$$

so that the item difficulty scale within each Stage h is set by restricting its Class 1 item difficulty parameters to be the same as those in Stage 1. Together, this system constitutes a necessary set of restrictions for identifying the model. An empirical check on the identification status of a Saltus model with a particular configuration of b's and a particular set of data is discussed in Section 3.3.

2.1.2 A Special Case

Wilson (1989) has discussed the case in which arrival in Stage h is signaled by a drop in the difficulty of items in item Class h, relative to items in all other classes. This difficulty shift is maintained in higher stages. This structure corresponds to a set of constraints among Saltus parameters:

$$\tau_{hk} = 0 \text{ if } h<k,$$

and

$$\tau_{hk} = \tau_{h'k} \text{ if both } h \geq k \text{ and } h' \geq k.$$

In this case there are only H-1 unique values for Saltus parameters, which for convenience may be called simply $\tau_2, \ldots, \tau_H$.

2.2 The Population Model

For estimation purposes, we assume a population in which the proportion of examinees in each developmental Stage h is π_h, with $0<\pi_h<1$. Denote by $\boldsymbol{\pi}$ the vector $(\pi_1, \ldots, \pi_H)$.

The density function of θ for Stage h is denoted $g_h(\theta)$. We shall discuss two special cases for g: a normal solution, wherein $g_h(\theta)$ is distributed as $N(\mu_h, \sigma_h)$, and a (nearly) nonparametric approximation, wherein each g_h is characterized as a histogram over a grid of prespecified points. The weight or density at point q for Stage h is denoted ω_{hq}. For generality, we use $\boldsymbol{\alpha}$ to denote population density parameters. In the normal solution, $\boldsymbol{\alpha} = (\mu_1, \sigma_1, \ldots, \mu_H, \sigma_H)$; in the nonparametric approximation, $\boldsymbol{\alpha} = (\omega_{hq})$.

3.0 Marginal Estimation of Structural Parameters

Assuming the Saltus item response model, (4) is the *conditional* probability of a response pattern **x**. Assuming further the population model described above, the *marginal* probability of **x**, or the probability of observing **x** from an examinee selected at random from the population, is given as

$$p(\mathbf{x}) = p(\mathbf{x} \mid \boldsymbol{\beta}, \mathbf{T}, \boldsymbol{\pi}, \boldsymbol{\alpha})$$

$$= \sum_h \pi_h \int P_h(\mathbf{x} \mid \theta, \boldsymbol{\beta}, \mathbf{T}) g_h(\theta \mid \boldsymbol{\alpha}) d\theta . \tag{5}$$

Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$ be the response matrix of a sample of N examinees to n test items. A realization of **X** induces the marginal likelihood function for $(\boldsymbol{\beta}, \mathbf{T}, \boldsymbol{\pi}, \boldsymbol{\alpha})$, as the product over examinees of factors like (5):

$$L(\mathbf{X} \mid \boldsymbol{\beta}, \mathbf{T}, \boldsymbol{\pi}, \boldsymbol{\alpha}) = \prod_i p(\mathbf{x}_i \mid \boldsymbol{\beta}, \mathbf{T}, \boldsymbol{\pi}, \boldsymbol{\alpha}). \tag{6}$$

We refer to $\boldsymbol{\beta}$, **T**, $\boldsymbol{\pi}$, and $\boldsymbol{\alpha}$ as the *structural* parameters of the problem. Their number remains constant irrespective of N. The *incidental* parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, whose numbers increase proportionally as N increases, have been eliminated by marginalizing over their respective distributions as in (5). MML estimation proceeds by finding the values of the structural parameters that maximize (6).

Equation (6) is an "incomplete data" likelihood function of the form addressed by Dempster, Laird, and Rubin (1977) . Estimating the structural parameters would be straightforward if values of θ and ϕ were observed from each examinee along with his or her response vector **x**; this would be a "complete data" problem. The EM algorithm maximizes the incomplete-data likelihood (6) iteratively. The E-step, or expectation step of each cycle, calculates the expectations of the sufficient statistics that the complete-data problem would require, conditional on the observed data and provisional estimates of the structural parameters. The M-step, or maximization step, solves what looks like a complete-data maximum likelihood problem using these conditional expectations of sufficient statistics. The resulting maxima for the structural parameters are improved estimates of the incomplete-data solution, and serve as input to the next E-step.

We employ the variation of the EM algorithm used by Bock and Aitkin (1981) to estimate item parameters, by Mislevy (1984, 1986) to estimate item parameters and population distribution parameters, and by Mislevy and Verhelst (1990) to estimate the parameters of mixtures of IRT models. Saltus is in fact a special case of the mixture models addressed by Mislevy and Verhelst. The integration that appears in (5) is approximated by summation over a fixed grid of points. The E-step calculates, for each examinee, the conditional probabilities of belonging to each stage, and, within each stage, the probabilities that θ takes the various grid-point values. The grid points play the role of weighted pseudo-data points in the M-step.

3.1 Solving the "Complete Data" Problem

This section gives the ML solution that would obtain if values of θ and ϕ were observed for each sampled respondent along with x. Among the N sampled examinees, some number $Q \leq N$ distinct values of θ will have been observed, say $\Theta_1, \ldots, \Theta_q, \ldots, \Theta_Q$. Now define the following statistics. I_{ihq} is an indicator variable that takes the value 1 if Examinee i is in Stage h and has proficiency Θq, and is zero otherwise. N_h is the number of examinees observed to be in Stage h:

$$N_h = \sum_i \phi_{ih} = \sum_i \sum_q I_{ihq}. \tag{7}$$

N_{hq} is the number of examinees in Stage h with $\theta = \Theta_q$:

$$N_{hq} = \sum_i I_{ihq}. \tag{8}$$

R_{jhq} is the number of examinees in Stage h with $\theta = \Theta_q$ who responded correctly to Item j:

$$R_{jhq} = \sum_i x_{ij} I_{ihq}. \tag{9}$$

The complete data likelihood for (β, T, π, α) induced by the observation of X, θ, and ϕ can be written as

$$L^*(\beta,T,\pi,\alpha \mid X,\theta,\phi) = \prod_h P(N_h \mid \pi) \prod_q P(N_{hq} \mid N_h,\alpha) \prod_j P(R_{jhq} \mid N_{hq},\beta,T),$$

whence the complete data log likelihood

$$\lambda^*(\beta,T,\pi,\alpha \mid X,\theta,\phi) = \sum_h N_h \log \pi_h \sum_q N_{hq} \log g_h(\Theta_q \mid \alpha) \times$$

$$\sum_j \sum_k b_{jk} \{R_{jhq} \log \Psi_{jhk}(\Theta_q) + (N_{hq} - R_{jhq}) \log[1 - \Psi_{jhk}(\Theta_q)]\}. \tag{10}$$

ML estimation for the complete data problem proceeds by solving the likelihood equations, which are obtained by setting to zero the first derivatives of (10) with respect to each element of (β, T, π, α).

For elements of $\boldsymbol{\pi}$, one must impose the constraint that $\Sigma\pi_h=1$. This can be accomplished with a Lagrangian multiplier (e.g., Mislevy, 1984, 369-370). One then obtains a closed form solution for the proportion of examinees in each stage:

$$\hat{\pi}_h = N_h/N. \tag{11}$$

For elements of $\boldsymbol{\alpha}$, the likelihood equations are

$$\frac{\partial\lambda^*}{\partial\boldsymbol{\alpha}} = \sum_h \sum_q N_{hq} \frac{\partial \log g_k(\Theta_q|\,\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}} = 0\,. \tag{12}$$

A nonparametric ML estimate of g_h, for example, estimates the density at each point Θ_q by the proportion of examinees from Stage q observed to have that proficiency:

$$\hat{\omega}_{hq} = N_{hq}/N_h. \tag{13}$$

If normal distributions are assumed, their means are estimated as

$$\hat{\mu}_h = N_h^{-1} \sum_q \Theta_q N_{hq}\,. \tag{14}$$

If each normal distribution can have a different variance, then

$$\hat{\sigma}_h^2 = N_h^{-1} \sum_q (\Theta_q-\mu_h)^2 N_{hq}\,; \tag{15}$$

if all are assumed to have the same variance, then

$$\hat{\sigma}^2 = N^{-1} \sum_h \sum_q (\Theta_q-\mu_h)^2 N_{hq}\,. \tag{16}$$

Even in the complete data problem, closed form solutions for $\boldsymbol{\beta}$ and $\mathbf{T}$ are not forthcoming. They can be estimated together without heavy calculation, however, using Newton steps for each element. From a provisional estimate z^0 of a generic element z, an improved estimate is obtained as

$$z^1 = z^0 - \left\{ \frac{\partial\lambda^*}{\partial z} \bigg|_{z=z^0} \right\} \left\{ \frac{\partial^2\lambda^*}{\partial z^2} \bigg|_{z=z^0} \right\}^{-1}.$$

For elements of $\boldsymbol{\beta}$, the constraint that $\Sigma\beta_j=0$ must be taken into account. Defining

$$\beta_n = -\sum_{j=1}^{n-1} \beta_j$$

we obtain the required first and second derivatives shown below. For Item j, for j=1, ... , n-1,

$$\frac{\partial\lambda^*}{\partial\beta_j} = \sum_q \sum_h \sum_k b_{jk}[N_{hq}\Psi_{jhk}(\Theta_q)-R_{jhq}] - b_{nk}[N_{hq}\Psi_{nhk}(\Theta_q)-R_{nhq}] \qquad (17)$$

and

$$\frac{\partial^2\lambda^*}{\partial\beta_j^2} = -\sum_q \sum_h N_{hq} \sum_k b_{jk}\Psi_{jhk}(\Theta_q)[1-\Psi_{jhk}(\Theta_q)] + b_{nk}\Psi_{nhk}(\Theta_q)[1-\Psi_{nhk}(\Theta_q)]\,. \qquad (18)$$

For Saltus parameter τ_{hk}, for h=2, ... , H and k=2, ... , H,

$$\frac{\partial\lambda^*}{\partial\tau_{hk}} = \sum_q \sum_j b_{jk}[R_{jhq}-N_{hq}\Psi_{jhk}(\Theta_q)] \qquad (19)$$

and

$$\frac{\partial^2\lambda^*}{\partial\tau_{hk}^2} = -\sum_q N_{hq} \sum_j b_{jk}\Psi_{jhk}(\Theta_q)[1-\Psi_{jhk}(\Theta_q)]\,. \qquad (20)$$

Note that the summations over j in (19) and (20), which include the factor b_{jk}, serve merely to pick up terms for only those items in item class k.

Solving the likelihood equations for $\boldsymbol{\beta}$ and $\mathbf{T}$ requires provisional estimates of each to calculate the Ψ_{jhk} terms that appear in (17) - (20). Once they are computed, a Newton step is taken for each element in $\boldsymbol{\beta}$ and $\mathbf{T}$ to provide improved estimates. These are used again to calculate improved estimates of the Ψs for the next Newton step. This procedure ignores the cross second derivatives among the elements of $\boldsymbol{\beta}$ and $\mathbf{T}$, but, from good starting values, converges rapidly nonetheless.

3.2 Solving the Incomplete Data Problem

We make the simplifying assumption that θ parameters can take only Q possible values, namely $\Theta_1, \ldots, \Theta_Q$. These values will play the role of the observed values Θ_q discussed in the preceding section. In any actual application of the Saltus model, neither the values of θ_i nor $\boldsymbol{\phi}_i$ are known, so neither will be the values of the indicator variables I_{ihq}. If the values of the structural parameters $\boldsymbol{\beta}$, $\mathbf{T}$, $\boldsymbol{\pi}$, and $\boldsymbol{\alpha}$ were known, however, it would be possible to calculate the expected values of the I_{ihq}s given $\mathbf{x}_i$s:

$$\tilde{I}_{ihq} = E(I_{ihq} | \mathbf{x}_i, \boldsymbol{\beta}, \mathbf{T}, \boldsymbol{\pi}, \boldsymbol{\alpha})$$

$$= \frac{\pi_h \, g_h(\Theta_q | \boldsymbol{\alpha}) \, P_h(\mathbf{x}_i | \Theta_q, \boldsymbol{\beta}, \mathbf{T})}{\sum_k \pi_k \sum_r g_k(\Theta_q | \boldsymbol{\alpha}) \; P_k(\mathbf{x}_i | \Theta_r, \boldsymbol{\beta}, \mathbf{T})} . \tag{21}$$

In the E-step of the EM approach to maximizing the marginal likelihood function (6), one evaluates (21) using provisional estimates of $\boldsymbol{\beta}$, $\mathbf{T}$, $\boldsymbol{\pi}$, and $\boldsymbol{\alpha}$. From these, one obtains expectations of the summary statistics defined in (7) - (9); call them $\tilde{N}_h$, $\tilde{N}_{hq}$, and $\tilde{R}_{jhq}$. Note that the Θ_q values play the role that observed θ values played in the complete data solution. Now, however, rather than observed counts of examinees at such a point, we have expected values of those counts.

In the M-step, one uses $\tilde{N}_h$, $\tilde{N}_{hq}$, and $\tilde{R}_{jhq}$ in place of their observed counterparts to solve facsimiles of the complete data likelihood equations via (11) - (20). Cycles of E- and M-steps are continued until successive changes are suitably small. Because the EM algorithm can be slow to converge, accelerating methods such as Ramsay's (1975) may be employed.

Equation (21) will be recognized as an application of Bayes theorem, giving the posterior probability that $\theta_i = \Theta_q$ and $\phi_{ih} = 1$ after observing $\mathbf{x}_i$. The normalizing constant in the denominator is an approximation of $p(\mathbf{x}_i)$ as given in (5). During the E-step, one may therefore accumulate the sum $-2\,\Sigma \log p(\mathbf{x}_i)$ to track the performance of improvement in fit over cycles, or to compare the fit of various values of structural parameters. For example, one can evaluate the impact of setting a particular Saltus parameter to zero, or compare a normal solution with equal variances in all stages against a solution that permits different variances.

3.3 Approximating the Information Matrix

Under the grid-point approximation described above, a method described by Louis (1982, Section 3.2) provides an approximation of the observed information matrix for MML estimates of the structural parameters in the Saltus model. For brevity, denote the parameter $(\boldsymbol{\beta}, \mathbf{T}, \boldsymbol{\pi}, \boldsymbol{\alpha})$ by $\boldsymbol{\eta}$. Louis' approximation is a sum over subjects of cross-products of expected complete-data log likelihood first derivatives:

$$I(\eta) \approx \sum_i \left[\sum_h \sum_q \frac{\partial \lambda^*(\eta \mid x_i, I_{ihq}=1)}{\partial \eta} \tilde{I}_{ihq} \right] \left[\sum_h \sum_q \frac{\partial \lambda^*(\eta \mid x_i, I_{ihq}=1)}{\partial \eta'} \tilde{I}_{ihq} \right].$$

The required terms for $\boldsymbol{\beta}$ and $\mathbf{T}$ are simplified versions of (17) and (19) respectively:

$$\frac{\partial \lambda^*(\eta \mid x_i, I_{ihq}=1)}{\partial \beta_j} = [\Psi_{jh}(\Theta_q) - x_{ij}] - [\Psi_{nh}(\Theta_q) - x_{ij}]$$

and

$$\frac{\partial \lambda^*(\eta \mid x_i, I_{ihq}=1)}{\partial \tau_{km}} = \sum_j b_{jm}[x_j - \Psi_{jkm}(\Theta_q)] .$$

Incorporating the constraint that the π's must sum to one, we obtain for π_h, for $h=1, \ldots, H-1$,

$$\frac{\partial \lambda^*(\eta \mid x_i, I_{ihq}=1)}{\partial \pi_h} = \pi_h^{-1} - \pi_H^{-1} .$$

For means and variances in the normal solution,

$$\frac{\partial \lambda^*(\eta \mid x_i, I_{ihq}=1)}{\partial \mu_h} = \frac{\Theta_q - \mu_h}{\sigma_h^2}$$

and

$$\frac{\partial \lambda^*(\eta \mid x_i, I_{ihq}=1)}{\partial \sigma_h^2} = \frac{(\Theta_q - \mu_h)^2 - \sigma_h^2}{2\sigma_h^4} .$$

If the observed information matrix is positive definite and the solution is the global maximum of the likelihood, its inverse is a large-sample approximation of the sampling variance of the MML estimates. In particular, square roots of the diagonal entries of I^{-1} are large-sample standard errors.

In addition to indicating the precision with which structural parameters have been estimated, the observed information matrix contributes to an understanding of the identification status of the model. As noted above, resolving the scale indeterminacies is necessary but not sufficient for identification. Another necessary condition is that the true information matrix be positive definite. Since the observed information matrix is a consistent estimate of the information matrix, a positive definite observed information matrix is supportive evidence of *local identification*. That is, in the neighborhood of the MML estimates, changes in parameter values imply changes in modelled response probabilities. The reader is referred to McHugh (1956) and Goodman (1974) for additional discussion of these issues in the closely-related context of latent class analysis.

3.4 Starting Values

The closer starting values are to final estimates, the fewer EM cycles will be required. Good starting values for the Saltus model can be based on Wilson's (1989) approximate estimation procedures. Modified slightly to conform to the identifying constraints specified in this presentation, the required steps are as follows.

1. Assign each examinee to a stage based on his observed response pattern. This will be straightforward in those cases in which successive stages imply greater probabilities of correct response to all items; total scores then identify "most likely" values of stage membership. In other cases, however, total scores will not suffice--as when moving to a higher stage means higher probabilities of success for some item classes, but lower probabilities for classes of items formerly answered correctly for the wrong reasons. Here provisional assignments for some examinees will depend on their relative successes in contrasting item classes. If it is still not possible to identify a most likely stage from among two or more possibilities, assign the examinee to one of them at random.

2. Use as initial estimates of π the proportions of examinees provisionally assigned to the stages. If no examinees have been assigned to a stage, use a small value such as .25/H as the starting value for that stage and adjust other probabilities accordingly.

3. Obtain estimates of item and person parameters under the simple Rasch model independently for each stage, using only the examinees provisionally assigned to that stage. If an item has a zero or perfect score, assign it a logit value based on Cohen's (1979) approximation for an item with a score of 1 or 1 less than the maximum score, respectively. Linearly transform the results so that
   a. the item parameter estimates for Stage 1 are centered at zero, and
   b. the average item difficulty for item Class 1 takes the same value in all stage calibrations.

4. Use as starting values for β the item parameter estimates from the Stage 1 calibration run.

5. To calculate starting values for $\boldsymbol{\alpha}$, use person ability estimates from each stage's calibration run, rescaled by the linear transformations applied to item difficulties applied in Step 3 above. For example, if normal distributions have been posited, calculate the mean and standard deviation of rescaled $\hat{\theta}$'s of the examinees provisionally assigned to each stage.

6. Calculate the average item difficulty in each item Class k in each rescaled calibration run h, denoting the results $\bar{\beta}_{hk}$. Use as starting values for $\mathbf{T}$ the values

$$\tau_{hk} = \bar{\beta}_{hk} - \bar{\beta}_{1k}\,, \quad h=2, ..., H;\ k=2, ..., H.$$

If additional constraints have been posited among τ's, appropriate averages or contrasts of the values so obtained may be used.

4.0 Empirical Bayes Estimates of Examinee Parameters

Once final estimates of structural parameters have been obtained, posterior probabilities of stage membership can be calculated for any examinee, and θ can be estimated conditional on stage membership. One begins by evaluating the expectations of the indicator variables I_{ihq} as shown in (21), using the MML estimates of $\boldsymbol{\beta}$, $\mathbf{T}$, $\boldsymbol{\pi}$, and $\boldsymbol{\alpha}$. For a response vector $\mathbf{x}_i$, the empirical Bayes approximation of probability of membership in Stage h is given as

$$P(\phi_{ih}=1 \mid \mathbf{x}_i) \approx \sum_q \tilde{I}_{ihq}\ . \tag{22}$$

Conditional on membership in Stage h, the posterior expectation of θ is approximated as

$$\bar{\theta}_{ih} = E(\theta \mid \phi_{ih}=1, \mathbf{x}_i) \approx \sum_q \Theta_q \tilde{I}_{ihq} / \sum_q \tilde{I}_{ihq}, \tag{23}$$

and the posterior variance is

$$\mathrm{Var}(\theta \mid \phi_{ih}=1, \mathbf{x}_i) \approx (\sum_q \Theta_q^2 \tilde{I}_{ihq} - \bar{\theta}_{ih}^2 \sum_q \tilde{I}_{ihq}) / \sum_q \tilde{I}_{ihq}\ . \tag{24}$$

5.0 Example 1: Simulated Data

This section describes a modest simulation comparing the performance of the MML algorithm with a solution treating examinees' stage memberships as if they were known true parameter values. Wilson's (1984) original approximations were based on a joint maximum likelihood (JML) estimation algorithm, and proceeded by first using an auxiliary algorithm to place each person into one or the other of the Saltus stages. This classification was not altered in the course of the algorithm. Under these circumstances, there is no mixture present, so the model is considerably simplified. The approach was found to give poor results under even generous conditions, and Wilson devised a correction based on "tailored simulations" to bring the estimates of the Saltus parameters closer to generating values. This was not a very satisfactory situation, and, in part, motivated this paper. In this simulation, we use an MML algorithm rather than a JML algorithm to estimate the remaining item and examinee-group parameters, to focus the comparison on the way examinee group membership is handled. In addition we judged that "tailored simulation", although somewhat efficacious in the previous work, should not be a part of the comparison. It is a complex and time-consuming process that few analysts would perform in practice.

Two-class Saltus item-response data were generated in a 2x2 design, based on the following two factors:

- The number of items in each Saltus class: moderate (10) or small (4). One would expect more difficulty recovering parameters with the smaller number of items, because less information is available about examinees' stage memberships.
- The value of the discontinuity parameter τ_{22}: moderate (1.5) or small (0.5). One would expect the smaller discontinuity value to cause more difficulty in parameter

recovery, again because classification of examinees according to stage membership is more problematic.

Each condition was replicated ten times, with 500 simulees drawn from each of two normally-distributed examinee stage groups, with means of -1.5 and 0.5 and standard deviations of .25. Saltus parameters were estimated for each replication under both the MML approach with a normal distribution and the "$\hat{\phi}$ as ϕ" approach.

Table 1 gives the generating values and the averages of the parameter estimates over the ten replications for the 10-items-per-class conditions, for both the moderate and small discontinuity conditions. There were ten items in each of two Saltus levels (items 1 to 10 and 11 to 20, respectively), with difficulties uniformly spread from -1.5 to 1.5.

Insert Table 1 about here

Consider first the combination of conditions that was expected to provide the best results, namely moderate number of items and moderate discontinuity. For the mixture model algorithm (column 3), the item parameters have been estimated quite well and the size of the Saltus stage groups is quite accurate, but the Saltus parameter has been underestimated by 0.11, or about 7 to 8 percent of its value. The ability distributions have been recaptured well. The "$\hat{\phi}$ as ϕ" approach (column 4), estimates item difficulties in the right order, but inflated away from zero. The Saltus parameter is overestimated by almost 300 percent, although the proportional representation of the Saltus stage groups is about right. The mean of the lower group is over a half a logit above its generating value, and its standard deviation is somewhat larger than it should be. The second stage's mean is well-estimated, and its standard deviation is also too large. Wilson's "tailored simulations" would have reduced the overestimation of the Saltus parameter, but would not have addressed any of the other problems.

The fifth column of Table 1 shows MML results for the small discontinuity condition. Compared to the moderate discontinuity condition, the item parameters are slightly deflated towards zero, and the size of the Stage 1 group has been estimated as .54 rather than .50. The Saltus parameter has again been underestimated, this time by 28 percent of its generating value. The stage means have both been overestimated somewhat, but their standard deviations behaved differently: the first is about twice as large as the generating value, while the second is only half as large. Column 6 contains the results for the "$\hat{\phi}$ as ϕ" approach. Here the item difficulties are inflated away from zero to about the same extent that the mixture model estimates were deflated back towards zero, and the size of Stage 1 group has been estimated as .56 rather than .50. Once again the Saltus parameter is greatly overestimated, this time by 500 percent. Both stage means have shrunk towards zero considerably, and both standard deviations are inflated, although to different degrees.

Table 2 presents generating values and results for the 4-items-per-class conditions. Among MML estimates (column 3), the item parameters have been estimated quite well and the size of the Saltus stage groups is quite accurate, but the Saltus parameter has again been underestimated, by about 10 percent. The ability distributions have been recaptured fairly well, although the standard deviation of the Stage 2 group is underestimated. The "$\hat{\phi}$ as ϕ" approach (column 4) shows an entirely different picture. The item difficulties are in the right order, but all are inflated away from zero somewhat. The Saltus parameter is overestimated by almost 200 percent, and the size of the Stage 1 group is overestimated. The mean of this lower group is almost logit above its generating value while the Stage 1 group's mean is less than it should be. Both standard deviations are overestimated.

Insert Table 2 about here

The fifth column of Table 2 shows the MML results for the small discontinuity condition. Compared to the moderate discontinuity condition, the item parameters have been deflated towards zero, and the size of Stage 1 group has been overestimated even more. The Saltus parameter has again been underestimated—essentially as zero. The stage group means have both been overestimated again, but their standard deviations have behaved differently: the first is about twice as large as the generating value, the second is about half as large. Column 6 contains the corresponding "$\hat{\phi}$ as ϕ" results. Here the item difficulties are slightly inflated away from zero, and the size of the Stage 1 group has been considerably overestimated. Once again the Saltus parameter is greatly overestimated, this time by 300 percent. Both stage group means have been reduced towards a common value, while both standard deviations are inflated.

In summary, the most salient of the results from the simulations are as follows:

1. Under the moderate number of items condition, and the moderate discontinuity condition, MML gives very good parameter recovery, with the exception of an underestimate of the Saltus parameter of an order somewhat less than 10 percent.
2. Under the mixed conditions (i.e., the "better" condition for one factor, and the "poorer" condition for the other), the mixture model gives good parameter recovery.
3. Under the small number of items condition and the small discontinuity condition, the mixture model condition gives a noticeably poorer estimation of several parameters, especially the Saltus parameter.
4. The "$\hat{\phi}$ as ϕ" approach gives uniformly poor estimates for the Saltus parameter, invariably overestimating it. The other parameters follow roughly the same relative patterns as for the MML results, although they are worse in almost all cases.

6.0 Example 2: Mixed Number Subtraction

The data analyzed in this example are responses of 325 junior high school students to 20 open-ended items dealing with mixed-number subtraction, gathered by Kikumi Tatsuoka and her colleagues. More detailed descriptions of the data and extensive cognitive analyses of the domain can be found in Klein, Birenbaum, Standiford, and Tatsuoka (1981), and an analysis based on Tatsuoka's "rule-space" approach appears in Tatsuoka (1990). We neglect many aspects of this rich data set in the following example, in order to illustrate how the Saltus model captures a key feature of in the domain: increasing competence possesses both qualitative and quantitative aspects, as learners master procedures and become more proficient in applying them. We contrast the Saltus solution with an analysis based on the RM shown as (1) and the 2-parameter logistic item response model:

$$P(x_j=1 \mid \theta, \alpha_j, \beta_j) = \Psi[\alpha_j(\theta-\beta_j)],$$

where α_j, the item slope parameter, indicates the sensitivity to which the probability of a correct response to item j reacts to changes in θ. Items with high values of α_j are considered to be good at discriminating high from low competence, from the perspective of the 2PL.

Table 3 presents the text of the items, percents-correct, and item parameter estimates under the RM and 2PL. These item parameters were obtained with Mislevy and Bock's (1989) *PC BILOG* program, assuming a normal distribution for θ and setting the scale so that the arithmetic mean of the estimated βs was 0 and the geometric mean of the αs was 1. Because we renumbered the items in order to group them in Saltus classes, the original Klein et al. item numbers are also shown. The item classes are based on whether an item requires two key procedures for its solution: finding a common denominator, and converting between mixed numbers and improper fractions. Items in Class 1 require neither; items in Class 2 require finding a common denominator; items in

Class 3 require converting, and possibly finding a common denominator as well. This implies that the qualitative aspect of students' is signaled by acquiring the common-denominator skill, then the converting skill. This path of development is not necessary either logically or psychologically, but it is not unreasonable to posit in this example because it accords with the instructional sequence.

Insert Table 3 about here

There is a clear pattern in the percentages of correct response. The items in each item class are of similar difficulty, and the average difficulties increase from the first class, to the second, to the third, with average percents correct of .73, .55, and .34. The RM difficulty parameters reflect this pattern directly, since they are nearly linear transformation of logits. The RM of the probabilities would suggest increasing competence to take the form of uniformly increasing chances of correct response on all items, in the logit metric. The 2PL would also posit linear increases in items' logits of correct response, but allow for faster or slower rates from one item to another, in proportion to their α parameters. Note the systematically higher 2PL slopes for the Class 2 and Class 3 items. The 2PL represents a substantially better fit to the actual response data, improving BILOG's chi-square index of comparative fit by 416 at the cost of 20 additional parameters (i.e., slopes).

Tables 4 through 6 present the results of the MML Saltus analysis, with normal distributions fitted within developmental stages. The Saltus solution offers a slightly greater improvement over the RM than does the 2PL—449 chi-square units at the cost of 12 additional parameters (4 τs, 3 means and standard deviations, and 2 independent proportions). The Saltus βs in Table 4 are item difficulty parameters for examinees in Stage 1. They are more spread out than those of the RM, indicating that for these examinees, exhibiting a large gap between the items in Class 1 and the items in Classes 2 and 3. The gap closes considerably when we look at the difficulty estimates that pertain to Stage 2 examinees; Class 2 items become just as easy for these students as Class 1 items. The shift is by the amount of the τ_{22} parameter in Table 5. Class 3 items still remain relatively difficult for Stage 2 examinees. The discontinuity associated with examinees in Stage 3 is the drop in difficulty of Class 3 items.

Insert Tables 4-6 about here

In addition to the shifts in relative item difficulties, the developmental stages are also distinguished in terms of their θ distributions (noting, of course, that θ has a different meaning for each stage, in terms of its implications for success on items from different classes). Figure 1 illustrates the relative locations of item difficulties and examinee distributions for the three stages. The locations of the Class 1 items set the scale; they are identical across the three panels. Being in Stage 1 typically implies middling chances of answering Class 1 items correctly, and practically no chance at Class 2 or 3 items. The Stage 2 line shows a noticeably higher θ distribution and a marked drop in the relative difficulty of Class 2 items. The Stage 3 line shows a slightly higher θ distribution and a marked drop in the relative difficulties of Class 3 items. These patterns are reflected in Table 7, which combines stage means with item parameters to give typical probabilities of correct response to each item from examinees of different classes.

Insert Figure 1 and Table 7 about here

Table 8 further details the discontinuities that Saltus can accomodate by showing observed responses and modeled probabilities for five examinees. We see that...

- *Examinee 4* got only half the items right, in a pattern spread across item classes. The RM and the 2PL accomodate this pattern well. Saltus handles it with a posterior concentrated on Stage 3, with a low θ value. There are enough Class 2 and Class 3 items correct to believe the student is beginning to use common denominator and converting procedures, but is not working with accuracy and consistency; this concords with missing two of the six easy Class 1 items.

- *Examinee 7* got half the Class 1 items right, three of the Class 2 items, and none of the Class 3 items. From the point of view of the RM and 2PL, some correct Class 3 responses would be expected. Saltus Stage 2 accords well with pattern, accounting for a dropoff between Class 2 and Class 3 items for students at this stage.
- *Examinee 12* got two Class 1 items right, one Class 2 item, and no Class 3 items. All models and all stages within Saltus agree in the predictions about the Class 1 items, but Saltus Stage 1 accords with this pattern best. For a student low in Class 1, correct answers to Class 2 and Class 3 items would be more rare than the RM or 2PL would predict.
- *Examinee 18* answered all Class 1 and Class 2 items correctly, but only three Class 3 items. This is a prototypical example of a Saltus Stage 2 pattern. For a student with this many correct responses, the RM and 2PL predict relatively fewer successes on Class 1 and 2 items, and relatively more successes on Class 3 items.
- *Examinee 536* also has Stage 2 as most probable stage under Saltus with a posterior probability of .67. There is an appreciable .33 probability for Stage 3, however, since half of the Class 3 items were answered correctly.

In this example, the improvements of fit over the Rasch model offered by both the 2PL and Saltus clearly indicate that there is more going on in the data than the RM can capture. The Saltus approach the potential role of theories about learning in the domain to provide inferences about the *nature* of students' competencies.

7.0 Conclusion

This paper has described a marginal maximum likelihood (MML) estimation algorithm for Wilson's (1984, 1989) Saltus model. The algorithm's performance was compared with that of joint maximum likelihood (JML), in which estimates of subjects' unobservable Saltus group memberships based on their total scores are treated as known. Substantial improvements were observed for tests of moderate length (10 items per class) and short length (4 items per class), in which misclassification of subjects is most likely to occur. Biases in estimates of structural parameters were eliminated almost competely for the moderate-length test, but not for the short test. In addition to reducing estimation biases, MML provides standard errors for item and Saltus parameter estimates that appropriately incorporate uncertainty due to imperfect information about examinees' Saltus group memberships.

References

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

Cohen, L. (1979). Approximate expressions for parameter estimates in the Rasch model. *British Journal of Mathematical and Statistical Psychology, 32*, 113-120.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society* (Series B), *39*, 1-38.

Embretson, S.E. (1985). Multicomponent latent trait models for test design. In S.E. Embretson (Ed.), *Test design: Developments in psychology and psychometrics*. Orlando: Academic Press.

Falmagne, J-C. (1989). A latent trait model via a stochastic learning theory for a knowledge space. *Psychometrika, 54*, 283-303.

Gagné, R.M. (1968). Learning hierarchies. *Educational Psychologist, 6*, 1-9.

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist, 36*, 923-936.

Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R. Ronning, J. Glover, J.C. Conoley, & J. Witt (Eds.), *The influence of cognitive psychology on testing and measurement: The Buros-Nebraska Symposium on measurement and testing* (Vol. 3). Hillsdale, NJ: Erlbaum.

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

Haertel, E.H. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement, 8*, 333-346.

Klein, M.F., Birenbaum, M., Standiford, S.N., & Tatsuoka, K.K. (1981). *Logical error analysis and construction of tests to diagnose student "bugs" in addition and* *subtraction of fractions*. Research Report 81-6. Urbana, IL: Computer-based Education Research Laboratory, University of Illinois.

Louis, T.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B, 44*, 226-233.

Masters, G., & Mislevy, R.J. (1993). New views of student learning: Implications for educational measurement. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests*. Hillsdale, NJ: Erlbaum.

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

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

Mislevy, R. J. (1986). Bayes model estimation in item response models. *Psychometrika, 51*, 177-195.

Mislevy, R.J., & Bock, R.D. (1989). *PC-BILOG 3: Item analysis and test scoring with binary logistic models*. Mooresville, IN: Scientific Software Inc.

Mislevy, R. J., and Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55*, 195-215.

Paulson, J.A. (1986). Latent class representation of systematic patterns in test responses. *Technical Report ONR-1*. Portland, OR: Psychology Department, Portland State University.

Ramsay, J. O. (1975). Solving implicit equations in psychometric data analysis. *Psychometrika, 40*, 361-372.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research/Chicago: University of Chicago Press (reprint).

Samejima, F. (1983). A latent trait model for differential strategies in cognitive strategies. *ONR Research Report 83-1*. Knoxville, TN: University of Tennessee.

Siegler, R.S. (1981). Developmental sequences within and between concepts. *Monograph of the Society for Research in Child Development, 46*.

Snow, R.E., & Lohman, D.F. (1989). Implications of cognitive psychology for educational measurement. In R.L. Linn (Ed.), *Educational measurement* (3rd Ed.) (pp. 263-331). New York: American Council on Education/Macmillan.

Tatsuoka, K.K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20*, 345-354.

Tatsuoka, K.K. (1990). Toward an integration of item response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M.G. Shafto, (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). Hillsdale, NJ: Erlbaum.

Thompson, P.W. (1982). Were lions to speak, we wouldn't understand. *Journal of Mathematical Behavior, 3*, 147-165.

Wilson, M. (1984). *A Psychometric model of hierarchical development*. Unpublished doctoral dissertation, University of Chicago.

Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin, 105*(2), 276-289.

Table 1
Generating Values and Estimates for the Moderate Number-of-Items Condition

Parameter	Generating Values	τ_{22}=1.5 Marginal Solution	τ_{22}=1.5 Solution treating $\hat{\phi}$ as ϕ	τ_{22}=0.5 Marginal Solution	τ_{22}=0.5 Solution treating $\hat{\phi}$ as ϕ
β_1	-1.50	-1.52	-2.25	-1.45	-1.89
β_2	-1.40	-1.37	-2.15	-1.38	-1.86
β_3	-1.30	-1.32	-2.11	-1.29	-1.79
β_4	-1.20	-1.20	-2.02	-1.16	-1.68
β_5	-1.10	-1.08	-1.92	-1.06	-1.60
β_6	-1.00	-0.98	-1.84	-0.87	-1.44
β_7	-0.90	-0.92	-1.78	-0.90	-1.47
β_8	-0.80	-0.74	-1.64	-0.74	-1.33
β_9	-0.60	-0.58	-1.51	-0.57	-1.20
β_{10}	-0.50	-0.43	-1.38	-0.42	-1.07
β_{11}	0.50	0.44	1.09	0.45	0.94
β_{12}	0.60	0.59	1.30	0.57	1.07
β_{13}	0.80	0.79	1.56	0.75	1.26
β_{14}	0.90	0.85	1.65	0.83	1.35
β_{15}	1.00	0.97	1.82	1.00	1.54
β_{16}	1.10	1.10	1.99	1.08	1.63
β_{17}	1.20	1.19	2.13	1.14	1.70
β_{18}	1.30	1.32	2.27	1.27	1.86
β_{19}	1.40	1.39	2.34	1.34	1.94
β_{20}	1.50	1.50	2.45	1.43	2.06
τ_{22}	-	1.39	4.37	0.36	2.44
π_1	0.50	0.50	0.51	0.54	0.56
π_2	0.50	0.50	0.49	0.46	0.44
μ_1	-1.50	-1.54	-0.91	-1.37	-0.80
μ_2	0.50	0.60	0.49	0.66	-0.27
σ_1	0.25	0.25	0.40	0.51	0.87
σ_2	0.25	0.21	0.43	0.13	0.45

Table 2

Generating Values and Estimates for the Small Number-of-Items Condition

Parameter	Generating Values	τ_{22}=1.5		τ_{22}=0.5	
		Marginal Solution	Solution treating $\hat{\phi}$ as ϕ	Marginal Solution	Solution treating $\hat{\phi}$ as ϕ
β_1	-1.50	-1.45	-1.72	-1.37	-1.64
β_2	-1.20	-1.19	-1.46	-1.07	-1.38
β_3	-1.00	-0.98	-1.27	-0.84	-1.17
β_4	-0.50	-0.45	-0.80	-0.29	-0.70
β_5	0.50	0.49	0.86	0.37	0.72
β_6	1.00	0.94	1.24	0.83	1.16
β_7	1.20	1.18	1.45	0.99	1.32
β_8	1.50	1.46	1.70	1.38	1.70
τ_{22}	-	1.38	2.95	-0.09	1.55
π_1	0.50	0.51	0.55	0.59	0.63
π_2	0.50	0.50	0.45	0.41	0.37
μ_1	-1.50	-1.46	-0.61	-1.21	-0.64
μ_2	0.50	0.58	-0.21	1.09	-0.06
σ_1	0.25	0.24	0.76	0.47	0.77
σ_2	0.25	0.10	0.48	0.08	0.39

Table 3

Item Text, Percents-Correct, and Saltus Difficulty Parameter Estimates

Item	Tatsuoka Item #	Text	Percent Correct	RM Difficulty	2PL Difficulty	2PL Slope
Saltus Class 1 Items						
1	6	$\frac{6}{7}-\frac{4}{7}=$	.79	-1.36	-1.46	.77
2	8	$\frac{2}{3}-\frac{2}{3}=$	.71	-.92	-1.23	.44
3	9	$3\frac{7}{8}-2=$	.69	-.86	-3.97	.12
4	12	$\frac{11}{8}-\frac{1}{8}=$	.71	-.94	-.97	.65
5	14	$3\frac{4}{5}-3\frac{2}{5}=$	.75	-1.16	-1.10	.85
6	16	$4\frac{5}{7}-1\frac{4}{7}=$	.74	-1.09	-1.05	.81
Saltus Class 2 Items						
7	1	$\frac{5}{3}-\frac{3}{4}=$	.50	-.04	.29	1.04
8	2	$\frac{3}{4}-\frac{3}{8}=$	.56	-.31	.06	1.68
9	3	$\frac{5}{6}-\frac{1}{9}=$	.51	-.05	.31	1.36
10	5	$4\frac{3}{5}-3\frac{4}{10}=$	.61	-.51	-.89	.27
Saltus Class 3 Items						
11	4	$3\frac{1}{2}-2\frac{3}{2}=$	.37	.54	.86	1.96
12	7	$3-2\frac{1}{5}=$	.33	.76	1.10	.98
13	10	$4\frac{4}{12}-2\frac{7}{12}=$	.31	.84	1.08	2.28
14	11	$4\frac{1}{3}-2\frac{4}{3}=$	.37	.56	.89	1.25
15	13	$3\frac{3}{8}-2\frac{5}{6}=$	.31	.82	1.10	4.58
16	15	$2-\frac{1}{3}=$	.38	.49	.84	1.08
17	17	$7\frac{3}{5}-\frac{4}{5}=$	.34	.69	1.02	1.15
18	18	$4\frac{1}{10}-2\frac{8}{10}=$	.41	.37	.73	1.03
19	19	$7-1\frac{4}{3}=$	.26	1.10	1.31	1.75
20	20	$4\frac{1}{3}-1\frac{5}{3}=$	.31	.84	1.11	1.61

Table 4

Saltus Item Parameter Estimates

Item	β	SE(β)	Implied Within-Stage Difficulty		
			Stage 1	Stage 2	Stage 3
Saltus Class 1 Items					
1	-2.94	.15	-2.94	-2.94	-2.94
2	-2.34	.14	-2.34	-2.34	-2.34
3	-2.26	.14	-2.26	-2.26	-2.26
4	-2.38	.14	-2.38	-2.38	-2.38
5	-2.66	.14	-2.66	-2.66	-2.66
6	-2.57	.14	-2.57	-2.57	-2.57
Saltus Class 2 Items					
7	0.00	.16	0.00	-2.85	-1.20
8	-0.52	.16	-0.52	-3.37	-1.73
9	-0.02	.16	-0.02	-2.88	-1.23
10	-0.94	.16	-0.94	-3.79	-2.14
Saltus Class 3 Items					
11	1.32	.18	1.32	0.32	-1.80
12	1.77	.18	1.77	0.77	-1.36
13	1.97	.18	1.97	0.96	-1.16
14	1.36	.18	1.36	0.35	-1.77
15	1.93	.18	1.93	0.93	-1.19
16	1.20	.18	1.20	0.20	-1.93
17	1.64	.18	1.64	0.64	-1.49
18	0.95	.18	0.95	-0.05	-2.18
19	2.51	.19	2.51	1.51	-0.62
20	1.97	.18	1.97	0.96	-1.16

Table 5

Saltus Parameter Estimates (Standard Errors in Parentheses)

Item Class	Examinee Stage 1	2	3
1	0.00*	0.00*	0.00*
2	0.00*	2.85 (0.20)	1.21 (0.13)
3	0.00*	1.00 (0.09)	3.13 (0.08)

* Fixed at zero for model identification.

Table 6

Saltus Examinee-Stage Estimates

Parameter	Stage 1	Stage 2	Stage 3
π	0.45	0.25	0.31
μ	-2.27	-0.77	-0.44
σ	0.68	0.90	0.85

Table 7

Modelled Average Percent-Correct for Saltus Classes

Item	Stage 1	Stage 2	Stage 3
Saltus Class 1 Items			
1	0.66	0.90	0.92
2	0.52	0.83	0.87
3	0.50	0.82	0.86
4	0.53	0.83	0.87
5	0.60	0.87	0.90
6	0.57	0.86	0.89
Average	**0.56**	**0.85**	**0.89**
Saltus Class 2 Items			
7	0.09	0.89	0.68
8	0.15	0.93	0.78
9	0.10	0.89	0.69
10	0.21	0.95	0.85
Average	**0.14**	**0.92**	**0.75**
Saltus Class 3 Items			
11	0.03	0.25	0.80
12	0.02	0.18	0.71
13	0.01	0.15	0.67
14	0.03	0.25	0.79
15	0.01	0.16	0.68
16	0.03	0.28	0.82
17	0.02	0.20	0.74
18	0.04	0.33	0.85
19	0.01	0.09	0.54
20	0.01	0.15	0.67
Average	**0.02**	**0.20**	**0.73**

Table 8

Posterior Distributions for Selected Subjects

	Posterior for θ			Observed Responses and Modeled Probabilities of Correct Response																			
Model	p	Mean	SD	Class 1 Items						Class 2 Items				Class 3 Items									
Examinee 4																							
Observed				1	1	0	0	1	1	0	1	1	0	1	1	0	0	0	1	0	0	0	0
RM	-	-.13	.30	.8	.7	.7	.7	.7	.7	.5	.5	.5	.6	.3	.3	.3	.3	.3	.3	.3	.4	.2	.3
2PL	-	.53	.24	.8	.7	.6	.7	.8	.8	.6	.7	.6	.6	.3	.4	.2	.4	.1	.4	.4	.4	.2	.3
Saltus																							
Stage 1	.07	-1.18	.47	.9	.8	.8	.8	.8	.8	.2	.3	.2	.4	.1	.1	.0	.1	.0	.1	.1	.1	.0	.0
Stage 2	.18	-1.23	.51	.9	.8	.7	.8	.8	.8	.8	.9	.8	.9	.2	.1	.1	.2	.1	.2	.1	.2	.1	.1
Stage 3	.75	-1.65	.41	.8	.7	.7	.7	.7	.7	.4	.5	.4	.6	.5	.4	.4	.5	.4	.6	.5	.6	.3	.4
Examinee 7																							
Observed				1	0	0	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
RM	-	-.59	.31	.7	.6	.6	.6	.6	.6	.4	.4	.4	.5	.2	.2	.2	.2	.2	.3	.2	.3	.2	.2
2PL	-	-.05	.30	.8	.7	.7	.7	.8	.7	.5	.6	.5	.6	.4	.3	.3	.4	.3	.4	.3	.4	.2	.3
Saltus																							
Stage 1	.25	-1.85	.48	.8	.6	.6	.6	.7	.7	.1	.2	.1	.3	.0	.0	.0	.0	.0	.1	.0	.1	.0	.0
Stage 2	.75	-2.00	.50	.7	.6	.6	.6	.7	.6	.7	.8	.7	.9	.1	.1	.1	.1	.1	.1	.1	.1	.0	.1
Stage 3	.00	-2.17	.42	.7	.5	.5	.6	.6	.6	.3	.4	.3	.5	.4	.3	.3	.4	.3	.4	.3	.5	.2	.3
Examinee 12																							
Observed				0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
RM	-	-1.12	.34	.6	.5	.4	.5	.5	.5	.3	.3	.3	.4	.2	.1	.1	.2	.1	.2	.1	.2	.1	.1
2PL	-	-.98	.40	.6	.5	.5	.5	.5	.5	.3	.3	.3	.4	.2	.1	.1	.2	.1	.2	.2	.2	.1	.1
Saltus																							
Stage 1	.98	-2.53	.47	.6	.5	.4	.5	.5	.5	.1	.1	.1	.2	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0
Stage 2	.02	-2.72	.47	.6	.4	.4	.4	.5	.5	.5	.7	.5	.7	.1	.0	.0	.0	.0	.1	.0	.1	.0	.0
Stage 3	.00	-2.69	.42	.6	.4	.4	.4	.5	.5	.2	.3	.2	.4	.3	.2	.2	.3	.2	.3	.2	.4	.1	.2

(continued)

Table 8, continued

Posterior Distributions for Selected Subjects

Model	Posterior for θ: p	Posterior for θ: Mean	Posterior for θ: SD	Class 1 Items						Class 2 Items				Class 3 Items									
Examinee 18																							
Observed				1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	1	0	0
RM	-	.45	.29	.9	.8	.8	.8	.8	.8	.6	.7	.6	.7	.5	.4	.4	.5	.4	.5	.4	.5	.3	.4
2PL	-	.64	.23	.9	.8	.8	.8	.9	.8	.7	.7	.7	.8	.5	.5	.4	.5	.5	.5	.5	.6	.4	.4
Saltus																							
Stage 1	.00	-.33	.46	.9	.9	.9	.9	.9	.9	.4	.6	.4	.7	.2	.1	.1	.2	.1	.2	.1	.2	.1	.1
Stage 2	.98	-.21	.50	.9	.9	.9	.9	.9	.9	.9	1.0	.9	1.0	.4	.3	.2	.4	.2	.4	.3	.5	.2	.2
Stage 3	.02	-.94	.44	.9	.8	.8	.8	.9	.8	.6	.7	.6	.8	.7	.6	.6	.7	.6	.7	.6	.8	.4	.6
Examinee 536																							
Observed				1	0	1	1	1	1	1	1	1	1	0	0	1	0	1	1	1	0	0	1
RM	-	.60	.30	.9	.8	.8	.8	.5	.8	.7	.7	.7	.8	.5	.5	.4	.5	.4	.5	.5	.6	.4	.4
2PL	-	1.12	.19	.9	.9	.9	.9	.9	.9	.8	.8	.8	.8	.6	.6	.6	.6	.6	.7	.6	.7	.5	.6
Saltus																							
Stage 1	.00	-.12	.45	.9	.9	.9	.9	.9	.9	.5	.6	.5	.7	.2	.1	.1	.2	.1	.2	.2	.3	.1	.1
Stage 2	.67	.05	.50	1.0	.9	.9	.9	.9	.9	1.0	1.0	1.0	1.0	.4	.3	.3	.4	.3	.5	.4	.5	.2	.3
Stage 3	.33	-.74	.45	.9	.8	.8	.8	.9	.9	.6	.7	.6	.8	.7	.7	.6	.7	.6	.8	.7	.8	.5	.6

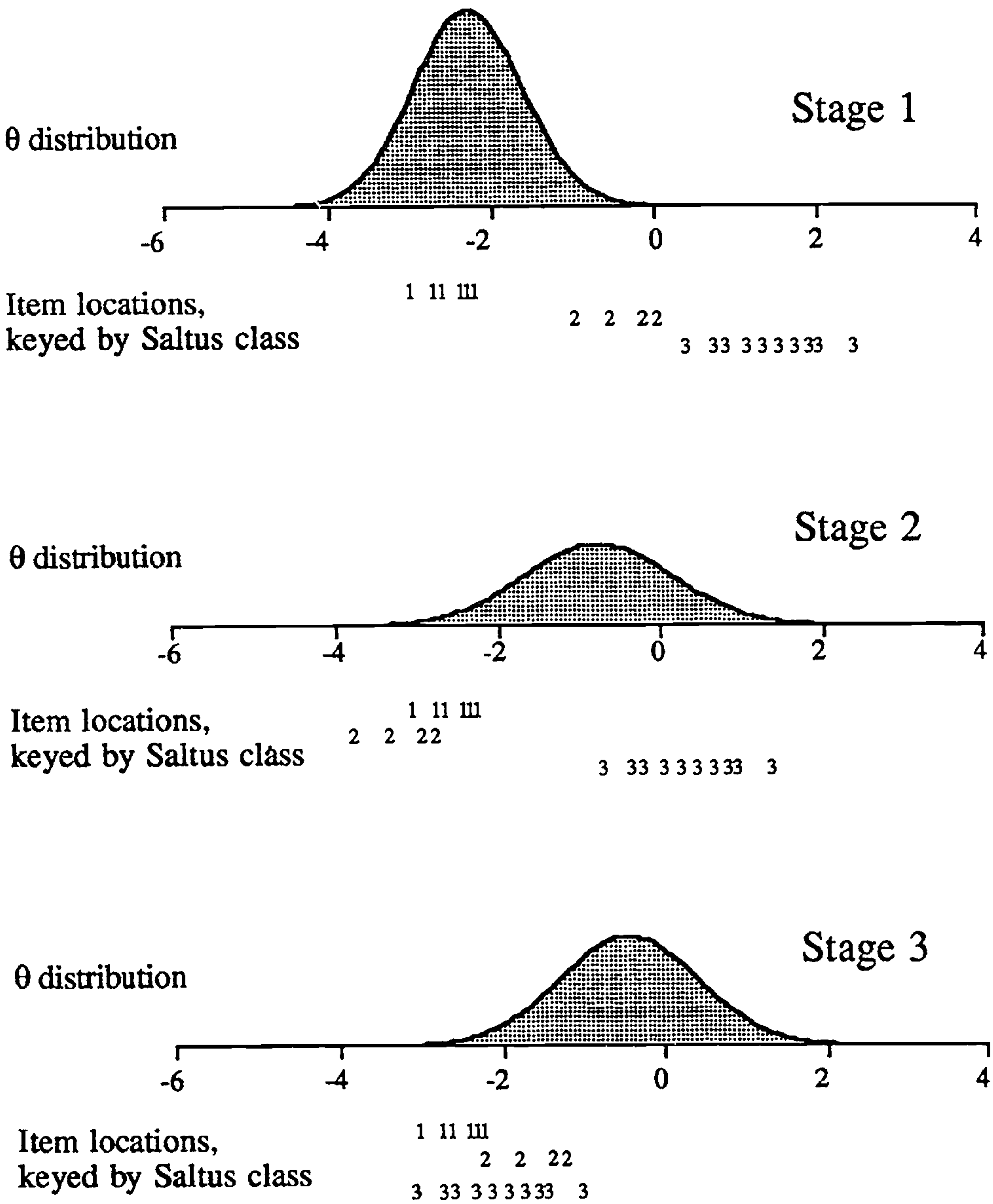

Figure 1

Modelled Saltus Item Locations and Class Membership Distributions

Dr. Terry Ackerman
Educational Psychology
260C Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Terry Allard
Code 1142CS
Office of Naval Research
800 N. Quincy St.
Arlington, VA 22217-5000

Dr. Nancy Allen
Educational Testing Service
Princeton, NJ 08541

Dr. Gregory Anrig
Educational Testing Service
Princeton, NJ 08541

Dr. Phipps Arabie
Graduate School of Management
Rutgers University
92 New Street
Newark, NJ 07102-1895

Dr. Isaac I. Bejar
Law School Admissions
Services
Box 40
Newtown, PA 18940-0040

Dr. William O. Berry
Director of Life and
Environmental Sciences
AFOSR/NL, N1, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Thomas G. Bever
Department of Psychology
University of Rochester
River Station
Rochester, NY 14627

Dr. Menucha Birenbaum
Educational Testing
Service
Princeton, NJ 08541

Dr. Bruce Bloxom
Defense Manpower Data Center
99 Pacific St.
Suite 155A
Monterey, CA 93943-3231

Dr. Gwyneth Boodoo
Educational Testing Service
Princeton, NJ 08541

Dr. Richard L. Branch
HQ, USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

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

Dr. David V. Budescu
Department of Psychology
University of Haifa
Mount Carmel, Haifa 31999
ISRAEL

Dr. Gregory Candell
CTB/MacMillan/McGraw-Hill
2500 Garden Road
Monterey, CA 93940

Dr. Paul R. Chatelier
Perceptronics
1911 North Ft. Myer Dr.
Suite 800
Arlington, VA 22209

Dr. Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy St.
Arlington, VA 22217-5000

Dr. Raymond E. Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB, TX 78235

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
Los Angeles, CA 90089-1061

Director
Life Sciences, Code 1142
Office of Naval Research
Arlington, VA 22217-5000

Commanding Officer
Naval Research Laboratory
Code 4827
Washington, DC 20375-5000

Dr. John M. Cornwell
Department of Psychology
I/O Psychology Program
Tulane University
New Orleans, LA 70118

Dr. William Crano
Department of Psychology
Texas A&M University
College Station, TX 77843

Dr. Linda Curran
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Timothy Davey
American College Testing Program
P.O. Box 168
Iowa City, IA 52243

Dr. Charles E. Davis
Educational Testing Service
Mail Stop 22-T
Princeton, NJ 08541

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Bldg., Rm. 1230F
University of Maryland
College Park, MD 20742

Dr. Sharon Derry
Florida State University
Department of Psychology
Tallahassee, FL 32306

Hei-Ki Dong
Bellcore
6 Corporate Pl.
RM: PYA-1K207
P.O. Box 1320
Piscataway, NJ 08855-1320

Dr. Neil Dorans
Educational Testing Service
Princeton, NJ 08541

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
(2 Copies)

Dr. Richard Duran
Graduate School of Education
University of California
Santa Barbara, CA 93106

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

Dr. George Engelhard, Jr.
Division of Educational Studies
Emory University
210 Fishburne Bldg.
Atlanta, GA 30322

ERIC Facility-Acquisitions
2440 Research Blvd., Suite 550
Rockville, MD 20850-3238

Dr. Marshall J. Farr
Farr-Sight Co.
2520 North Vernon Street
Arlington, VA 22207

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

Dr. Richard L. Ferguson
American College Testing
P.O. Box 168
Iowa City, IA 52243

Dr. Gerhard Fischer
Liebiggasse 5
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
U.S. Army Headquarters
DAPE-HR
The Pentagon
Washington, DC 20310-0300

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Chair, Department of
Computer Science
George Mason University
Fairfax, VA 22030

Dr. Robert D. Gibbons
University of Illinois at Chicago
NPI 909A, M/C 913
912 South Wood Street
Chicago, IL 60612

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. Susan R. Goldman
Peabody College, Box 45
Vanderbilt University
Nashville, TN 37203

Dr. Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque, NM 87131

Dr. Joseph McLachlan
Navy Personnel Research
and Development Center
Code 14
San Diego, CA 92152-6800

Alan Mead
c/o Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Timothy Miller
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. Ivo Molenaar
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The NETHERLANDS

Dr. E. Muraki
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Ratna Nandakumar
Educational Studies
Willard Hall, Room 213E
University of Delaware
Newark, DE 19716

Academic Progs. & Research Branch
Naval Technical Training Command
Code N-62
NAS Memphis (75)
Millington, TN 30854

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Norman, OK 73071

Head, Personnel Systems Department
NPRDC (Code 12)
San Diego, CA 92152-6800

Director
Training Systems Department
NPRDC (Code 14)
San Diego, CA 92152-6800

Library, NPRDC
Code 041
San Diego, CA 92152-6800

Librarian
Naval Center for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington, DC 20375-5000

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

Special Assistant for Research
Management
Chief of Naval Personnel (PERS-01JT)
Department of the Navy
Washington, DC 20350-2000

Dr. Judith Orasanu
Mail Stop 239-1
NASA Ames Research Center
Moffett Field, CA 94035

Dr. Peter J. Pashley
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

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

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Dr. Peter Pirolli
School of Education
University of California
Berkeley, CA 94720

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

Mr. Steve Reise
Department of Psychology
University of California
Riverside, CA 92521

Mr. Louis Roussos
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Donald Rubin
Statistics Department
Science Center, Room 608
1 Oxford Street
Harvard University
Cambridge, MA 02138

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37966-0900

Dr. Mary Schratz
4100 Parkside
Carlsbad, CA 92008

Mr. Robert Semmes
N218 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis, MN 55455-0344

Dr. Valerie L. Shalin
Department of Industrial
Engineering
State University of New York
342 Lawrence D. Bell Hall
Buffalo, NY 14260

Mr. Richard J. Shavelson
Graduate School of Education
University of California
Santa Barbara, CA 93106

Ms. Kathleen Sheehan
Educational Testing Service
Princeton, NJ 08541

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

Dr. Randall Shumaker
Naval Research Laboratory
Code 5500
4555 Overlook Avenue, S.W.
Washington, DC 20375-5000

Dr. Judy Spray
ACT
P.O. Box 168
Iowa City, IA 52243

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. David Thissen
Psychometric Laboratory
CB# 3270, Davie Hall
University of North Carolina
Chapel Hill, NC 27599-3270

Mr. Thomas J. Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row, Suite 501
Memphis, TN 38131

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

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

Elizabeth Wald
Office of Naval Technology
Code 227
800 North Quincy Street
Arlington, VA 22217-5000

Dr. Michael T. Waller
University of
Wisconsin-Milwaukee
Educational Psychology Dept.
Box 413
Milwaukee, WI 53201

Dr. Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. Thomas A. Warm
FAA Academy
P.O. Box 25082
Oklahoma City, OK 73125

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Douglas Wetzel
Code 15
Navy Personnel R&D Center
San Diego, CA 92152-6800

German Military
Representative
Personalstammamt
Koelner Str. 262
D-5000 Koeln 90
WEST GERMANY

Dr. Sherrie Gott
AFHRL/MOMJ
Brooks AFB, TX 78235-5601

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

Prof. Edward Haertel
School of Education
Stanford University
Stanford, CA 94305-3096

Dr. Ronald K. Hambleton
University of Massachusetts
Laboratory of Psychometric
and Evaluative Research
Hills South, Room 152
Amherst, MA 01003

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Patrick R. Harrison
Computer Science Department
U.S. Naval Academy
Annapolis, MD 21402-5002

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 13
San Diego, CA 92152-6800

Dr. Thomas M. Hirsch
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Paul W. Holland
Educational Testing Service, 21-T
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Ms. Julia S. Hough
Cambridge University Press
40 West 20th Street
New York, NY 10011

Dr. William Howell
Chief Scientist
AFHRL/CA
Brooks AFB, TX 78235-5601

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

Dr. Martin J. Ippel
Center for the Study of
Education and Instruction
Leiden University
P. O. Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr. Robert Jannarone
Elec. and Computer Eng. Dept.
University of South Carolina
Columbia, SC 29208

Dr. Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign, IL 61820

Professor Douglas H. Jones
Graduate School of Management
Rutgers, The State University
of New Jersey
Newark, NJ 07102

Dr. Brian Junker
Carnegie-Mellon University
Department of Statistics
Pittsburgh, PA 15213

Dr. Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh, PA 15213

Dr. J. L. Kaiwi
Code 442/JK
Naval Ocean Systems Center
San Diego, CA 92152-5000

Dr. Michael Kaplan
Office of Basic Research
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Jeremy Kilpatrick
Department of
Mathematics Education
105 Aderhold Hall
University of Georgia
Athens, GA 30602

Ms. Hae-Rim Kim
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State
University
Murfreesboro, TN 37132

Dr. Sung-Hoon Kim
KEDI
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

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
Box 7246, Meas. and Eval. Ctr.
University of Texas-Austin
Austin, TX 78703

Dr. James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

Dr. Patrick Kyllonen
AFHRL/MOEL
Brooks AFB, TX 78235

Ms. Carolyn Laney
1515 Spencerville Rod
Spencerville, MD 20868

Richard Lanterman
Commandant (G-PWP)
US Coast Guard
2100 Second St., SW
Washington, DC 20593-0001

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
1310 South Sixth Street
University of IL at
Urbana-Champaign
Champaign, IL 61820-6990

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541-0001

Mr. Hsin-hung Li
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Library
Naval Training Systems Center
12350 Research Parkway
Orlando, FL 32826-3224

Dr. Marcia C. Linn
Graduate School
of Education, EMST
Tolman Hall
University of California
Berkeley, CA 94720

Dr. Robert L. Linn
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Logicon Inc. (Attn: Library)
Tactical and Training Systems
Division
P.O. Box 85158
San Diego, CA 92138-5158

Dr. Richard Luecht
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. George B. Macready
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Evans Mandes
George Mason University
4400 University Drive
Fairfax, VA 22030

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

Dr. James R. McBride
HumRRO
6430 Elmhurst Drive
San Diego, CA 92120

Mr. Christopher McCusker
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. Robert McKinley
Educational Testing Service
Princeton, NJ 08541

Dr. David Wiley
School of Education
and Social Policy
Northwestern University
Evanston, IL 60208

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

Dr. Mark Wilson
School of Education
University of California
Berkeley, CA 94720

Dr. Eugene Winograd
Department of Psychology
Emory University
Atlanta, GA 30322

Dr. Martin F. Wiskoff
PERSEREC
99 Pacific St., Suite 4556
Monterey, CA 93940

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
03-0T
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Ms. Duanli Yan
Educational Testing Service
Princeton, NJ 08541

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

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550