DOCUMENT RESUME

ED 303 474 TM 012 463

AUTHOR	Mislevy, Robert J.; Verhelst, Norman
TITLE	Modeling Item Responses When Different Subjects Employ Different Solution Strategies.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Office of Naval Research, Washington, D.C. Psychological Sciences Div.
REPORT NO	ETS-RR-87-47-ONR
PUB DATE	Oct 87
CONTRACT	N00014-85-K-0683
NOTE	58p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Guessing (Tests); *Latent Trait Theory; *Maximum Likelihood Statistics; *Models; *Problem Solving; Spatial Ability
IDENTIFIERS	*Item Parameters; *Response Patterns; Spatial Tests

ABSTRACT

A model is presented for item responses when different examinees use different strategies to arrive at their answers and when only those answers, not choice or strategy or subtask results, can be observed. Using substantive theory to differentiate the likelihoods of response vectors under a fixed set of solution strategies, responses are modeled in terms of item parameters associated with each strategy, proportions of the population employing each, and the distributions of examinee parameters within each. Posterior distributions can then be obtained for each examinee, giving the probabilities that they employed each of the strategies and their proficiency under each. The ideas are illustrated with a conceptual example about response strategies for spatial rotation items, and a numerical example resolving a population of examinees into subpopulations of valid responders and random guessers. Four data tables and four graphs are presented. A 26-item list of references is provided. (Author/TJH)

RR-87-47-ONR

MODELING ITEM RESPONSES WHEN DIFFERENT SUBJECTS EMPLOY DIFFERENT SOLUTION STRATEGIES

Robert J. Mislevy

and

Norman Verhelst

CITO
(National Institute for Educational Measurement)
Arnhem, The Netherlands

This research was sponsored in part by the
Cognitive Science Program
Psychological Sciences Division
Office of Naval Research, under
Contract No. N00014-85-K-0683

Contract Authority Identification No.
NR 150-539

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, New Jersey

October 1987

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) RR-87-47-ONR	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Educational Testing Service 6b OFFICE SYMBOL *(If applicable)*	7a NAME OF MONITORING ORGANIZATION Cognitive Science Program, Office of Naval Research (Code 1142CS) 800 North Quincy Street
6c. ADDRESS *(City, State, and ZIP Code)* Princeton, NJ 08541	7b ADDRESS *(City, State, and ZIP Code)* Arlington, VA 22217-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION 8b OFFICE SYMBOL *(If applicable)*	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-85-K-0683
8c. ADDRESS *(City, State, and ZIP Code)*	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO 61153N; PROJECT NO RR04204; TASK NO RR04204-01; WORK UNIT ACCESSION NO NR 150-539

11 TITLE *(Include Security Classification)*

Modeling Item Responses When Different Subjects Employ Different Solution Strategies (Unclassified)

12 PERSONAL AUTHOR(S)

Robert J. Mislevy and Norman Verhelst

13a. TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT *(Year, Month, Day)*	15. PAGE COUNT
Technical	FROM ______ TO ______	October 1987	45

16. SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
FIELD	GROUP	SUB-GROUP	Differential strategies; Linear logistic test model; Item response theory; Mixture models
05	10		

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

A model is presented for item responses when different examinees employ different strategies to arrive at their answers, and when only those answers, not choic of strategy or subtask results, can be observed. Using substanti.e theory to differentiate the likelihoods of response vectors under a fixed set of solution strategies, we model responses in terms of item parameters associated with each strategy, proportions of the population employing each, and the distributions of examinee parameters within each. Posterior distributions can then be obtained for each examinee, giving the probabilities that they employed each of the strategies and their proficiency under each. The ideas are illustrated with a conceptual example about response strategies for spatial rotation items, and a numerical example resolving a population of examinees into subpopulations of valid responders and random guessers.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. James Lester	22b. TELEPHONE *(Include Area Code)* 202-696-4503 22c. OFFICE SYMBOL ONR 1142CS

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Modeling Item Responses When Different Subjects

Employ Different Solution Strategies[1]

Robert J. Mislevy

Educational Testing Service

and

Norman Verhelst

CITO
(National Institute for Educational Measurement)
Arnhem, The Netherlands

October 1987

[1]The first author's work was supported by Contract No. N00014-85-K-0683, Project Designation No. NR 150-539, from the Cognitive Science Program, Psychological Sciences Division, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government. We are grateful to Isaac Bejar, Neil Dorans, Norman Frederiksen, and Marilyn Wingersky for their comments and suggestions.

Modeling Item Responses When Different Subjects

Employ Different Solution Strategies

Abstract

A model is presented for item responses when different examinees employ different strategies to arrive at their answers, and when only those answers, not choice of strategy or subtask results, can be observed. Using substantive theory to differentiate the likelihoods of response vectors under a fixed set of solution strategies, we model responses in terms of item parameters associated with each strategy, proportions of the population employing each, and the distributions of examinee parameters within each. Posterior distributions can then be obtained for each examinee, giving the probabilities that they employed each of the strategies and their proficiency under each. The ideas are illustrated with a conceptual example about response strategies for spatial rotation items, and a numerical example resolving a population of examinees into subpopulations of valid responders and random guessers.

Key Words: Differential strategies
Item response theory
Linear logistic test model
Mixture models

Introduction

The standard models of item response theory (IRT), such as the 1-, 2-, and 3-parameter normal and logistic models, characterize examinees in terms of their propensities to make correct responses. Consequently, examinee parameter estimates are strongly related to simple percent-correct scores (adjusted for the average item difficulties, if not all examinees have been presented the same items). Item parameters characterize the regression of a correct response on this overall propensity toward correctness.

These models lend themselves well to tests in which all examinees employ the same strategy to solve the items. Comparisons among estimates of examinees' ability parameters are meaningful comparisons of their degrees of success in implementing the strategy. Item parameters reflect the number or complexity of the operations needed to solve a given item (Fischer, 1973).

The same models can prove less satisfactory when different examinees employ different strategies. The validity of using scores that convey little more than percent-correct to compare examinees who have used different strategies must first be called into question. And item parameters keyed only to a generalized propensity toward correctness will not reveal how a particular kind of item might be easy for examinees who follow one line of attack, but difficult for those who follow another.

Extensions of IRT to multiple strategies have several potential uses. In psychology, such a model would provide a rigorous analytic framework for testing alternative theories about cognitive processing (e.g., Carter, Pazak, and Kail, 1983). In education, estimates of how students solve problems could be more valuable than how many they solve, for the purposes of diagnosis, remediation, and curriculum revision (Messick, 1984). And even when a standard IRT model would provide reasonable summaries and meaningful comparisons for most examinees, an extended model allowing for departures along predetermined lines (e.g., malingering) would reduce estimation biases for the parameters in the standard model.

In contrast to standard IRT models, and, for that matter, to the "true score" models of classical test theory, a model that accommodates alternative strategies must begin with explicit statements about the processes by which examinees arrive at their answers. For example, items may be characterized in terms of the nature, number, and complexity of the operations required for their solution under each strategy that is posited.

The recent psychometric literature contains a few implementations of these ideas. Tatsuoka (1983) has studied performance on mathematics items in terms of the application of correct and incorrect rules, locating response vectors in a two-dimensional space based on an ability parameter from a standard IRT model and an index of lack of fit from that model. Paulson

(1985), analyzing similar data but with fewer rules, uses latent class models to relate the probability of correct responses on an item to the features it exhibits and the rules that examinees might be following. Yamamoto (1987) combines aspects of both of these models, positing subpopulations of IRT respondents and of non-scalable respondents associated with particular expected response patterns. Samejima's (1983) and Embretson's (1985) models for alternative strategies are expressed in terms of subtasks whose results are observed, in addition to the overall correctness or incorrectness of the item.

The present paper describes a family of multiple-strategy IRT models that apply when each examinee belongs to one of a number of exhaustive and mutually-exclusive classes that correspond to an item-solving strategy, and the responses from all examinees in a given class are in accordance with a standard IRT model. It is further assumed that for each item, its parameters under the IRT model for each strategy class can be related to known features of the item through psychological or pedagogical theory.

The next section of the paper gives a general description of the model. It is followed by a conceptual example that illustrates the key ideas. A two-stage estimation procedure is then presented. The first stage estimates structural parameters: basic parameters for test items, examinee population distributions, and proportions of examinees following each strategy The second stage estimates posterior distributions for individual examinees: the probability that they belong to each strategy class and the conditional distribution of their ability for each class. A numerical example resolves examinees into classes of valid responders and random guessers. The final section discusses some implications of the approach for educational and psychological testing.

The Response Model

This section lays out the basic structure for a mixture of constrained item response models. Discussion will be limited to dichotomous items for notational convenience, but the extensions to polytomous and continuous observations are straightforward.

We begin by briefly reviewing the general form of an IRT model. The probability of response x_{ij} (1 if correct, 0 if not) from person i to item j is given by an IRT model as

$$p(x_{ij}|\theta_i,\beta_j) = [f(\theta_i,\beta_j)]^{x_{ij}}[1-f(\theta_i,\beta_j)]^{1-x_{ij}} \qquad (1)$$

where θ_i and β_j are real (and possibly vector-valued) parameters associated with person i and item j respectively, and f is a known, twice-differentiable, function whose range is the unit interval. Under the usual IRT assumption of local independence, the conditional probability of the response pattern $\mathbf{x}_i = (x_{i1},\ldots,x_{in})$ of person i to n items is the product of n expressions like (1):

$$p(x_i|\theta_i,\beta) = \prod_{j=1}^{n} p(x_{ij}|\theta_i,\beta_j).$$

It may possible to express item parameters as functions of some smaller number of more basic parameters $\alpha = (\alpha_1,\ldots,\alpha_M)$ that reflect the effects of M salient characteristics of items; i.e., $\beta_j = \beta_j(\alpha)$. An important example of this type is the Linear Logistic Test Model (LLTM; Fischer, 1973, Schieblechner, 1972). Under the LLTM, the item response function is the one-parameter logistic (Rasch) model, or

$$p[x_{ij}|\theta_i,\beta_j(\alpha)] = \exp[x_{ij}(\theta_i-\beta_j)]/[1+\exp(\theta_i-\beta_j)],$$

and the model for item parameters is linear:

$$\beta_j(\alpha) = \sum_{m=1}^{M} Q_{jm}\,\alpha_m = Q_j'\,\alpha \quad .$$

The elements of α are contributions to item difficulty associated with the M characteristics of items, presumably related to the number or nature of processes required to solve them. The elements of the known vector Q_j indicate the extent to which item j exhibits each characteristic. Fischer (1973), for example, models the difficulty of the items in a calculus test in terms of the number of times an item requires the application of each of seven differentiation rules. Q_{jm} is the number of times that rule m must be employed in order to solve Item j.

Consider now a set of items that may be answered by means of K different strategies. It need not be the case that all are equally effective, nor even that all generally lead to correct responses. Not all strategies need be available to all examinees. We make the following assumptions.

1. Each examinee is applying the same one of these strategies for all the items in the set. (In the final section, we discuss prospects for relaxing this assumption to allow for strategy-switching).

2. The responses of an examinee are observed but the strategy he or she has employed is not.

3. The responses of examinees following Strategy k conform to an item response model of a known form.

4. Substantive theory posits relationships between observable features of items and the probabilities of success enjoyed by members of each strategy class. The relationships may be known either fully or only partially (as when the Q matrices in LLTM-type models are known but the basic parameters are not).

Let the k'th element in the K-dimensional vector ϕ_i take the value one if examinee i follows Strategy k, and zero if not. Extending the notation introduced above, we may write the conditional probability of response pattern $\mathbf{x}_i$ as

$$p(x_i|\phi_i,\theta_i,\alpha) = \prod_k\{\prod_j[f_k(\theta_{ik},\beta_{jk})]^{x_{ij}}[1-f_k(\theta_{ik},\beta_{jk})]^{1-x_{ij}}\}^{\phi_{ik}} \quad (2)$$

where $\beta_{jk} \equiv \beta_{jk}(\alpha)$ gives the item parameter(s) for Item j under Strategy k.

It will be natural in certain applications to partition basic parameters for items in accordance with strategy classes; that is, $\alpha = (\alpha_1,\ldots,\alpha_K)$. When there are K versions of the LLTM, for example, differences among strategies are incorporated into the model by K different vectors Q_{jk}, k=1,...,K, that relate Item j to each of the strategies:

$$\beta_{jk} = \sum_m Q_{jkm}\,\alpha_{km} = Q'_{jk}\,\alpha_k \ .$$

The item difficulty parameter for Item j under Strategy k, then, is a weighted sum of elements in α_k, the basic parameter vector associated with Strategy k; the weights Q_{jkm} indicate the degree to which each of the features m, as relevent under Strategy k, are present in Item j. This situation will be illustrated in the following example.

Example 1: Alternative strategies for spatial tasks

The items of certain tests intended to measure spatial visualization abilities admit to solution by nonspatial analytic strategies (French, 1965; Kyllonen, Lohman, and Snow, 1984; Pelligrino, Mumaw, and Shute, 1985). Consider items in which subjects are shown a drawing of a three-dimensional target object, and asked whether a stimulus drawing could be the same object after rotation in the plane of the picture. In addition to rotation, one or more key features of the stimulus may differ from the those of target. A subject may solve the item either by rotating the target mentally the required degree and recognizing the match (Strategy 1), or by employing analytic reasoning to detect feature matches without performing rotation (Strategy 2).

Consider further a hypothetical three-item test comprised of such items. Each item will be characterized by (1) rotational displacement, of 60, 120, or 180 degrees, and by (2) the number of features that must be matched. Table 1 gives the features of the items in the hypothetical test.

Insert Table 1 about here

Each subject i will be characterized by two vectors. In the first, $\phi_i = (\phi_{i1},\phi_{i2})$, ϕ_{ik} takes the value 1 if Subject i employs

Strategy k and 0 if not. In the second, $\theta_i = (\theta_{i1}, \theta_{i2})$, θ_{ik} characterizes the proficiency of Subject i if he employs Strategy k. Only one of the elements of θ_i is involved in producing Subject i's responses, but we do not know which one.

Suppose that for subjects employing a rotational strategy, probability of success is given by the one-parameter logistic (Rasch) item response model:

$$p(x_{ij}|\theta_{i1},\beta_{j1},\phi_1=1) = \exp[x_{ij}(\theta_{i1}-\beta_{j1})]/[1+\exp(\theta_{i1}-\beta_{j1})] .$$

Here θ_{i1} is the proficiency of Subject i at solving tasks by means of the rotational strategy, and β_{j1} is the difficulty of Item j under the rotational strategy.

It is usually found that the time required to solve mental rotation tasks is linearly related to rotational displacement. To an approximation, so are log-odds of success (Tapley and Bryden, 1977). We assume that under the rotational strategy, item parameters take the following form:

$$\beta_{j1} = Q_{j11}\,\alpha_{11} + \alpha_{12} ,$$

where Q_{j11} encodes the rotational displacement of Item j--1 for 60 degrees, 2 for 120 degrees, and 3 for 180 degrees--and α_{11} is the incremental increase in difficulty for each increment in rotation; and α_{12} is a constant term, for which a coefficient $Q_{j12}=1$ is implied for all items. If $\alpha_{11}=1$ and $\alpha_{12}=-2$, the item parameters

β_{j1} that are in effect under Strategy 1 are as shown in the second column of Table 2.

Insert Table 2 about here

A Rasch model will also be assumed for subjects employing Strategy 2, the analytic strategy, but here the item parameters depend on the number of features that must be matched:

$$\beta_{j2} = Q_{j21}\, \alpha_{21} + \alpha_{22} \ ,$$

where Q_{j21} is the number of salient features, α_{21} is the incremental contribution to item difficulty of an additional feature, α_{22} is a constant term, and Q_{j22}=1 implicitly for all items. If α_{21}=1.5 and α_{22}=-2.5, we obtain the item parameters that are in effect under Strategy 2. They appear in the third column of Table 2.

Note that the items have been constructed so that items that are relatively hard under one strategy are easy under the other. Strategy choice cannot be inferred from observed response patterns unless patterns are more likely under some strategies and less likely under others.

The response pattern 011, for example, has a correct answer to an item that is easy under the Strategy 2 but hard under Strategy 1, and an incorrect answer to an item that is hard under

Strategy 2 but easy under Strategy 1. Figure 1 plots the likelihood function for the response vector 011 under both strategies; that is, $p[x=(011)|\theta_k,\phi_k=1]$ for $k=1,2$ as a function of θ_1 and θ_2 respectively. The maximum of the likelihood under Strategy 2 is about eight times as high as the maximum attained under Strategy 1.

Insert Figure 1 about here

We can make probabilistic statements about individual subjects if we know the proportions of people who choose each strategy, or $\pi_k = p(\phi_k=1)$, and the distributions of proficiency of those using each strategy class, or $g_k(\theta_k) = p(\theta_k|\phi_k=1)$. Suppose that (i) θ_1 and θ_2 both follow standard normal distributions among the subjects that have chosen to follow them, and (ii) three times as many subjects follow Strategy 1 as follow Strategy 2--i.e., $\pi_1 = 3/4$ and $\pi_2 = 1/4$. This joint prior distribution is pictured in Figure 2.

Insert Figure 2 about here

Routine application of Bayes theorem then yields the joint posterior density function for ϕ and $\theta_k|\phi_k=1$ for $k=1,\ldots,K$:

$$p(\theta_k=\theta,\phi_k=1|x,\pi,\alpha) \propto p[x|\phi_k=1,\theta,\beta_k(\alpha)]\ \pi_k\ g_k(\theta)\ , \qquad (3)$$

where

$$p[x|\phi_k=1,\theta,\beta_k(\alpha)] = \prod_j \exp\{x_{ij}[\theta-\beta_{jk}(\alpha)]\}/\{1+[\theta-\beta_{jk}(\alpha)]\} .$$

The reciprocal of the constant of proportionality required to normalize (3) is the marginalization of the right side, or

$$\sum_k \pi_k \int p[x|\phi_k=1,\theta,\beta_k(\alpha)]\ g_k(\theta)\ d\theta .$$

The posterior distribution induced by (011) is shown in Figure 3. Marginalizing with respect to θ_k amounts to summing the area under the curve for Strategy k, and gives the posterior probability that $\phi_k=1$--that is, that the subject has employed Strategy k. The resulting values for this response pattern are $P(\phi_1=1|x=011)=.28$ and $P(\phi_2=1|x=011)=.72$. The prior probabilities favoring Strategy 1 have been revised substantially in favor of Strategy 2. The conditional posterior for θ_1 given $\phi_1=1$ has a mean and standard deviation of about .32 and .80. Corresponding values for the distribution of θ_2 given $\phi_2=1$ are .50 and .81.

Insert Figure 3 about here

Parameter Estimation

This section discusses estimation procedures for mixtures of IRT models. A two-stage procedure is described. The first stage integrates over θ and ϕ distributions to obtain a so-called marginal likelihood function for the structural parameters of the problem--the basic parameters for items, the proportions of subjects employing each strategy, and the parameters of the θ distributions of subjects employing each strategies. Maximum likelihood estimates are obtained by maximizing this likelihood function. If preferred, Bayes modal estimates can be obtained by similar numerical procedures by multiplying the likelihood by prior distributions for the structural parameters. The second stage takes the resulting point estimates of structural parameters as known, and calculates aspects of the posterior distribution of an individual examinee--e.g., $p(\phi_k=1|x)$ and $p(\theta_k|\phi_k=1,x)$.

Stage 1: Estimates of Structural Parameters

Equation 2 gives the conditional probability of the response vector x given θ and ϕ, or $p(x|\theta,\phi,\alpha)$. Consider a population in which strategies are employed in proportions π_k and within-strategy proficiencies have densities $g_k(\theta_k|\eta_k)$ among the examinees using them. The marginal probability of x for an examinee selected at random from this population is

$$p(x|\alpha,\pi,\eta) = \sum_k \pi_k \int p(x|\theta_k,\phi_k=1,\alpha)\, g_k(\theta_k|\eta_k)\, d\theta_k \, . \qquad (4)$$

For brevity, let ξ denote the extended vector of all structural parameters, namely (α,π,η). The loglikelihood for ξ induced by the observation of the response vectors $X = (x_1,\ldots,x_N)$ of N subjects is a constant plus the sum of the logs of terms like (4) for each subject:

$$\lambda = \sum_{i=1}^{N} \log p(x_i|\xi)$$

$$= \sum_i \sum_k \phi_{ik} \log \int p[x_i|\theta_k,\phi_k=1,\beta_k(\alpha)]\ g_k(\theta_k|\eta_k)\ d\theta_k$$

$$+ \sum_i \sum_k \phi_{ik} \log \pi_k \ . \qquad (5)$$

Let S be the vector of first derivatives, and H the matrix of second derivatives, of λ with respect to ξ. Under regularity conditions, the maximum likelihood estimates $\hat{\xi}$ solve the likelihood equation S=0, and a large-sample approximation of the matrix of estimation errors is given by the negative inverse of H evaluated at $\hat{\xi}$.

A standard numerical approach to solving likelihood equations is to use some variation of Newton's method. Newton-Raphson iterations, for example, improve a provisional estimate ξ^0 by adding the correction term $-H^{-1} S \Big|_{\xi=\xi^0}$. Fletcher-Powell iterations avoid computing and inverting H by using an approximation of H^{-1} that is built up from changes in S from one cycle to the next.

These solutions have the advantage of rapid convergence if starting values are reasonable--often fewer than 10 iterations are necessary. S and H can be difficult to work out, however, and all parameters must be usually be dealt with simultaneously because the off-diagonal elements in H needn't be zero. For these reasons, a computationally simpler but slower-converging solution based on Dempster, Laird, and Rubin's (1977) EM algorithm will now be described as well. The approximation uses discrete representations for the g_ks, so the relatively simple "finite mixtures" case obtains (Dempster, Laird, and Rubin, 1977)

Suppose that for each k, subject proficiency under Strategy k can take only the L(k) values $\theta_{k1},\ldots,\theta_{kL(k)}$. The density g_k is thus characterized by these points of support and by the weights associated with each, $g_k(\theta_{k\ell}|\eta_k)$. Define the subject variable $\psi_i = (\psi_{i11},\ldots,\psi_{iKL(K)})$, a vector of length Σ_k L(k) where the element $\psi_{ik\ell}$ is 1 if the proficiency of Subject i under Strategy k is $\theta_{k\ell}$ and 0 if not. There are a total of K 1s in ψ_i, one for each strategy--though again, only only of them is involved in producing x_i--the one associated with the strategy that Subject i happens to employ. Summations replace integrations in the loglikelihood, which can now be written as

$$\lambda = \sum_i \sum_k \phi_{ik} \sum_\ell \psi_{ik\ell} \log p[x_i|\theta_k=\theta_k,\phi_k=1,\beta_k(\alpha)]$$

$$+ \sum_i \sum_k \phi_{ik} \sum_\ell \psi_{ik\ell}\, g_k(\theta_{k\ell}|\eta_k)$$

$$+ \sum_i \sum_k \phi_{ik} \log \pi_k \,. \qquad (6)$$

If values of ϕ and ψ were observed along with values of x, ML estimation of ξ from (6) would be simpler. The basic parameter α appears only in the first term on the right side of (6), so that maximizing with respect to α need address that term only. When α consists of distinct subvectors for each strategy, even these subvectors lead to distinct maximization problems of lower order. The subpopulation parameters η appear in only the second term, separating them in ML estimation; they too lead to even smaller separate subproblems if η consists of distinct subvectors for each strategy. The population proportions π appear in only the last term. Unless they are further constrained, their ML estimates are simply observed proportions. The values of θ may be either specified a priori (as in Mislevy, 1986) or estimated from the data (as in de Leeuw and Verhelst, 1986). In the latter case, their likelihood equations have contributions from both the first and second terms, but the equations for the points of support under Strategy k involve data from only those subjects using Strategy k. Their cross second derivatives with points corresponding to other strategies are zero, although their cross derivatives with elements of α and η that are involved with the same strategy need not be.

The M-step of an EM solution requires solving a maximization problem of exactly this type, with one exception: the unobserved values of each ϕ_i and ψ_i are replaced by their conditional expections given x_i and provisional estimates of ξ, say ξ^0. The

E-step calculates these conditional expectations as follows. Denote by $l_{ik\ell}$ the following term in the marginal likelihood associated with Subject i, Strategy k, and proficiency value $\theta_{k\ell}$ within Strategy k:

$$l_{ik\ell} = p[x_i | \theta_k = \theta_{k\ell}, \phi_k = 1, \beta_k(\alpha)] \; g_k(\theta_k | \eta_k) \; \pi_k \quad .$$

The required conditional expectations are obtained as

$$\psi^0_{ik\ell} = E(\psi_{ik\ell} | x_i, \xi = \xi^0)$$

$$= l^0_{ik\ell} \Big/ \sum_{\ell'} l^0_{ik\ell'} \qquad (7)$$

and

$$\phi^0_{ik} = E(\phi_{ik} | x_i, \xi = \xi^0)$$

$$= l^0_{ik\ell} \Big/ \sum_{k'} \sum_{\ell'} l^0_{ik'\ell'} \quad . \qquad (8)$$

The EM formulation makes it clear how each subject contributes to the estimation of the parameters in all strategy classes, even though it is assumed that only one of them was relevant to the production of his responses. His data contribute to estimation for each strategy class is in the proportion to the probability that that strategy was the one he employed, given his observed response pattern.

In addition to its simplicity, the EM solution has the advantage of being able to proceed from even very poor starting values. The slowness with which it converges can be a serious drawback, however. Its rate of convergence depends on how well **x** determines examinees' θ and ϕ values. Accelerating procedures such as those described by Ramsay (1975) and Louis (1982) can be used to hasten convergence.

Stage 2: Posteriors for Individual Examinees

When the population parameters ξ are accurately estimated, the posterior density of the parameters of examinee i is approximately

$$p(\theta_{ik}=\theta,\phi_{ik}=1|x_i,\hat{\xi}) \propto p[x_i|\phi_k=1,\theta,\beta_k(\hat{\alpha})]\ \hat{\pi}_k\ g_k(\theta|\hat{\eta}_k) ,$$

where the reciprocal of the normalizing constant is obtained by first integrating the expression on the right over θ within each k, then summing over k. The posterior probability that Subject i used Strategy k is approximated by

$$P(\phi_{ik}=1|x_i,\hat{\xi}) = \int p(\theta_{ik}=\theta,\phi_{ik}=1|x_i,\hat{\xi})\ d\theta .$$

The examinee's posterior mean and variance for a given strategy class, given that that was the strategy employed, are approximated by

$$\tilde{\theta}_{ik} = \int \theta \; p(\theta_{ik}=\theta, \phi_{ik}=1 | x_i, \hat{\xi}) \; d\theta \; / \; P(\phi_{ik}=1 | x_i, \hat{\xi})$$

and

$$\tilde{\sigma}^2_{ik} = \int (\theta-\tilde{\theta}_{ik})^2 p(\theta_{ik}=\theta, \phi_{ik}=1 | x_i, \hat{\xi}) \; d\theta \; / \; P(\phi_{ik}=1 | x_i, \hat{\xi}) \; .$$

If the discrete approximation has been employed, (7) and (8) apply.

Example 2: A Mixture of Valid Responders and Random Guessers

Given appropriate instructions, examinees will omit multiple-choice test items when they don't know the answers rather than guess at random. The Rasch model may provide a good fit to such data if omits are treated as incorrect. If a small percentage of examinees responds at random to all items, however, their responses will bias the estimation of the item parameters that pertain to the majority of the examinees.

We may posit a two-class model, under which an examinee responds either in accordance with the Rasch model or guesses totally at random. For examinees in the latter class, probabilities of correct response are constant, e.g., at the reciprocal of the number of response alternatives to each item.

Using the procedures described in the preceding sections, it is possible to free estimates of the item parameters that pertain to the valid responders from biases due to random guessers, even though it is not known with certainty who the guessers are.

A mixture model for the (marginal) probability of response pattern x in this situation is

$$P(x_i|\xi) = \sum_{k=1}^{2} P(x_i|\phi_k=1,\xi)\ \pi_k ,$$

where Strategy Class 1 is the Rasch model and Class 2 is random guessing. The composition of ξ is now described. It includes first the strategy proportions π_1 and π_2. For the Rasch class, the basic parameters α_1 are item difficulty parameters b_j for $j=1,\ldots,n$. Suppose the distribution g_1 of proficiencies of subjects following the Rasch model is discrete, with L points of support $\theta = (\theta_1,\ldots,\theta_L)$ and associated weights $\omega = (\omega_1,\ldots,\omega_L)$. The (marginal) probability of response pattern x under Strategy 1 is

$$P(x|\phi_1=1,\alpha_1,\theta,\omega) = \sum_{\ell} \omega_\ell \prod_j \exp[x_j(\theta_\ell-b_j)]/[1+\exp(\theta_\ell-b_j)] .$$

Under the random guessing strategy, the basic parameters α_2 are the probabilities c_j of responding correctly to each item j. All subjects following this strategy are assumed to have the same probabilities of correct response, so no distribution g_2 enters the picture. For such subjects, the probability of response pattern **x** is simply

$$P(\mathbf{x}|\phi_2=1,\alpha_2) = \Pi\, c_j^{\,x_j} (1-c_j)^{1-x_j} \quad .$$

An artificial dataset was created for four items under this model in accordance with the following specifications. Of 1200 simulees in all, 1000 followed the Rasch model and 200 were random guessers, implying π_1=.833 and π_2=.167. The Rasch item parameters were $\alpha_1 \equiv (b_1,\ldots,b_4) = (-.511,-.105,.182,.405)$. A discrete density with six points of support was used to create the data for the Rasch class. The points and their corresponding proportions were as follows:

Point	Proportion
-1.204	.08
-.357	.17
.095	.25
.262	.25
.470	.17
.642	.08

The rates of correct response for the random guessers on the four items were $\alpha_2 \equiv (c_1,\ldots,c_4) = (.30, .35, .20, .15)$. The probability of each of the sixteen possible response patterns was calculated within each class, multiplied by the number of simulees in that class, summed over classes, and rounded to the nearest integer. The resulting data are shown in Table 3.

Insert Table 3 about here

A standard Rasch model was first fit to the data using the two-step marginal maximum likelihood procedures described by de Leeuw and Verhelst (1986). Conditional maximum likelihood (CML) estimates were first obtained for item parameters. Setting their scale by centering them around zero like the true item parameters for the Rasch class, the resulting values were (-.324, -.053, .127, .252). Note that these values are biased toward their center; the presence of random guessers blurs the distinctions among the differences in item difficulties. A three-point discrete distribution--the greatest number of points leading to an identified model for a four-item test--was next estimated for subjects. The expected counts of response patterns under this model are also shown in Table 3. A chi-square of 7.16 with 8 degrees of freedom results, indicating an acceptable fit for a sample of the size we have employed.

A mixture model of the generating form was then fit to the data, with two exceptions. First, the multiplicative form of the Rasch model was employed during calculations. Since maximum likelihood estimates are invariant under transformations, the estimates of the structural parameters obtained under the multiplicative form need merely be transformed back to the usual additive form shown above. Second, a three-point discrete distribution was again employed for the Rasch class, with the lowest point fixed at zero in the multiplicative scale. This corresponds to $\theta_1 = -\infty$ in the additive scale, implying incorrect responses to all items with probability one. (As it turns out, the estimated weight associated with this point will be zero.) The total number of parameters to be estimated, then, was 13:

- 2 free points in the Rasch distribution: θ_2 and θ_3.
- 2 free values for weights at the three points in the Rasch distribution: ω_1, ω_2, and ω_3, where $\Sigma\, \omega_\ell = 1$.
- 4 item parameters for the Rasch class: $\alpha_1=(b_1,\ldots,b_4)$.
- 4 item parameters for the guessing class: $\alpha_2=(c_1,\ldots,c_4)$.
- 1 relative proportion for class representation: π_2.

In light of the fact that only 15 degrees of freedom are available from the data, in the form of 16 response patterns whose counts that must sum to 1200, an unaccelerated EM solution converged painfully slowly. Fletcher-Powell iterations were employed instead, and they converged rapidly. The Rasch-only estimates described above were used as starting values for the Rasch class item parameters and population distribution. For the c's, a common value midway among the true values was used. For π_2, starting values of .10, .15, and .20 were used in three different runs. All runs converged to the same solution:

$\alpha_1 = (-.501, -.091, .193, .398);$

$\theta = (-\infty, -.534, .354);$

$\omega = (<10^{-10}, .319, .681);$

$\alpha_2 = (.287, .230, .179, .139);$

$\pi_2 = .164.$

Although the c's are slightly underestimated, the structure of the data has been reconstructed quite well. The expected counts of response patterns are also shown in Table 3. As they should, they yield a nearly perfect fit: a chi-square of .008 on 3 degrees of freedom. The improvement in chi-square is dramatic if not significant--it would be for larger samples or longer tests--but the removal of the bias in the Rasch item parameter estimates is the point of the exercise.

Table 4 shows conditional likelihoods of each response pattern given that an examinee is a guesser, a member of the Rasch class with $\theta = -.534$, and a member of the Rasch class with $\theta = .354$. The estimated proportions of the population in these categories are .164, .267, and .569 respectively. Multiplying these population probabilities times a pattern's corresponding likelihood terms, then normalizing, gives the posterior probabilities that also appear in the table. Posterior probabilities are given for membership in the guessing class, and for $\theta = -.534$ and $\theta = .354$ given membership in the Rasch class.

Insert Table 4 about here

Recall from the description of the EM solution that the data from an examinee is effectively distributed among strategy classes to estimate the item parameters within that class. This means that the responses of all examinees play a role in both estimating both b's and c's--but with weights in proportion to the posterior probabilities shown in Table 4. From responses to only four items, we never have overwhelming evidence that a particular examinee is a guesser. Only those with all incorrect responses can be judged more likely than not to have guessed. Had only those respondents been treated as guessers--and that would be the Bayesian modal estimate of strategy class--estimated c's would all have been zero. But employing a proportion of data from all patterns, even those with all items correct, yields estimated c's that essentially recover the generating values.

As a consequence of using the Rasch model for Strategy 1, the conditional posterior distributions given that a subject belongs to this class, or $p(\theta|x,\phi_1=1)$, are identical for all response patterns x with the same total score. The probability that an examinee belongs to the Rasch class vary considerably within patterns with the same score, however. For any given response pattern, the posterior probability of being in the Rasch class can be inferred from Table 4 as $1 - P(\phi_2=1|\mathbf{x})$. For patterns with exactly one correct response, these probabilities are, for Items 1-4 in turn, .869, .800, .687, and .519.

Discussion

Theories about the processes by which examinees attempt to solve test items play no role in standard applications of test theory, including conventional item response theory. Only a data matrix of correct and incorrect responses is addressed, and items and examinees are parameterized strictly on the basis of propensities toward correct response. When all that is desired is a simple comparison of examinees in terms of a general propensity of this nature, IRT models suffice and in fact offer many advantages over classical true-score test theory.

Situations for which standard IRT models prove less satisfactory involve a desire either to better understand the cognitive processes that underlie item response, or to employ theories about such processes to provide more precise or more valid measurement. Extensions of item response theory in this direction are exemplified by the Linear Logistic Test Model (Schieblechner, 1972; Fischer, 1973), Embretson's (1985) multicomponent models, Samejima'a (1983) model for multiple strategies, and Tatsuoka's (1983) "rule space" analyses.

The approach offered in this paper concerns situations in which different persons may choose different strategies from a number of known alternatives, but overall proficiencies provide meaningful comparisons among persons employing the same strategy. We suppose that strategy choice is not directly observed but can be inferred (with uncertainty) from response patterns on theoretical bases. Assuming that substantive theory allow us to differentiate our expectations about response patterns under different strategies, and that a subject applies the same strategy on all items, it is possible to estimate the parameters of IRT models for each strategy. It is further possible to calculate the probabilities that a given subject has employed each of the alternative strategies, and estimate his proficiency under each given that that was the one he used.

Assuming that a subject uses the same strategy on all items is obviously undesirable for many important problems. In a technical sense, the approach can be extended to allow for strategy-switching by defining additional strategy classes that are in effect combinations of different strategies for different items. Based on Just and Carpenter's (1985) finding that subjects sometimes apply whichever strategy is easier for a given problem, we might define three strategy classes for items like those in our Example 1:

- Always apply the rotational strategy;
- Always apply the analytic strategy;
- Apply whichever strategy is better suited to an item.

If items were constructed to run from easy to hard under the rotational strategy and hard to easy under the analytic, subjects using the third "mixed" strategy would find them easy, then harder, then easier again.

There are limitations to how far these ideas can be pressed in applications with binary data. Our second example showed that the misspecified Rasch model fit a four-item test acceptably well with a sample of 1200 subjects; in one way or another, more information would be needed to attain a sharper distinction between strategy classes and, correspondingly, more power to differentiate among competing models for the data. One source of information is more binary items. Fifty items rather than four, including some that are very hard under the Rasch strategy, would do. A different source of information available in other settings would be to draw from richer observational possibilities. Examples would include response latencies as well as correctness, eye-fixation patterns, and choices of incorrect alternatives that are differentially likely under different strategies.

Differentiating the likelihood of response patterns under different strategies is the key to successful applications of the approach. Its use would be recommended when identifying strategy classes is of primary importance to the selection or placement decision that must be made and overall proficiency is of secondary importance. The items in the test must then be constructed to maximize strategy differences, e.g., using items that are hard under one strategy but easy under another. Most tests in current use with standard test theory are not constructed with this purpose in mind; indeed, they are constructed so as to minimize differentiation among strategies, since it lowers the reliability of overall-propensity scores. When strategy class decisions are of interest, a conventional tests is not likely to provide useful information. (Although a battery of conventional tests might; differences in score profiles are analogous to differential likelihoods of item response patterns, but at a higher level of aggregation.)

In addition to the applications used in the preceding examples, a number of other current topics in educational and psychological research are amenable to expression in terms of mixtures of IRT models. We conclude by mentioning three.

Hierarchical development. Wilson's (1984, 1985) "saltus" model (Latin for "leap") extends the Rasch model to developmental patterns in which capabilities increase in discrete stages, by including stage parameters as well as abilities for persons, and stage parameters as well as difficulties for items. Examples would include Piaget's (1960) innate developmental stages and Gagne's (1962) learned acquisition of rules. Suppose that K stages are ordered in terms of increasing and cumulative competence. In our notation, ϕ would indicate the stage membership of a subject. In the highest stage, item responses follow a Rasch model with parameters b_j. Rasch models fit lower stages as well, but the item parameters are offset by amounts that depend on which stage the item can first be solved. Our basic parameters α would correspond to the item parameters for the highest stage and the offset parameters for particular item types at particular lower stages. Figure 4 gives a simple illustration in which items associated with higher stages have an additional increment of difficulty for subjects at lower stages. In applications such as Siegler's (1981) balance beam tasks, subjects at selected lower stages tend to answer certain types of higher-stage items correctly for the wrong reasons. In these cases, the offset works to give easier item difficulty parameters to those items in those stages.

Insert Figure 4 about here

Mental models for problem solving. In the introduction to their experimental study on mental models for electricity, Gentner and Gentner (1983) state

> Analogical comparisons with simple or familiar systems often occur in people's descriptions of complex systems, sometimes as explicit analogical models, and sometimes as implicit analogies, in which the person seems to borrow structure from the base domain without knowing it. Phrases like "current being routed along a conductor" and "stopping the flow" of electricity are examples (p. 99).

Mental models are important as a pedagogical device and as a guide to problem-solving. Inferring which models a person is using, based on a knowledge of how conceivable analogues help or hinder the solution of certain types of problems, provides a guide to subsequent training. In Gentner and Gentner's experiment, the problems concerned simple electrical circuits with series and parallel combinations of resistors and batteries. Popular analogies for electricity are flowing waters (Strategy 1) and "teeming crowds" of people entering a stadium through a few narrow turnstiles (Strategy 2). The water flow analogy facilitates battery problems, but does not help with resistor problems; indeed, it suggests an incorrect solution for the current in circuits with parallel resistors. The teeming crowd analogy facilitates problems on the combination of resistors, but is not informative about combinations of batteries. If a Rasch model holds for items within strategies, Gentner and Gentner's hypotheses correspond to constraints on the order of item difficulties with the two strategics. If each item type were replicatated enough times, it would be possible to make inferences about which model a particular examinee was using, in order to plan subsequent instruction.

Changes in intelligence over age. An important topic in the field of human development is whether, and how, intelligence changes as people age (Birren, Cunningham, and Yamamoto, 1983). Macrae (n.d.) identifies a weakness of most studies that employ psychometric tests to measure aging effects: total scores fail to reflect important differences in the strategies different subjects bring to bear on the items they are presented. Total score differences among age and educational-background groups on Raven's matrices test were not significant in the study she reports. But analyses of subjects' introspective reports on how they solved items revealed that those with academically oriented background were much more likely to have used the preferred "algorithmic" strategy over a "holistic" strategy than those with vocationally oriented backgrounds. Since the use of algorithmic strategies was found to increase probabilities of success differentially on distinct item types, this study would be amenable to IRT mixture modeling. Inferences could then be drawn about problem-solving approaches without resorting to more expensive and possibly unreliable introspective evidence.

References

Birren, J.E., Cunningham, W.R., and Yamamoto, K. (1983). Psychology of adult development and aging. Annual Review of Psychology, 34, 543-575.

Carter, P., Pazak, B., and Kail, R. (1983). Algorithms for processing spatial information. Journal of Experimental Child Psychology, 36, 284-304.

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.

Fischer, G.H. (1973). The linear logistic test model as an instrument in educational research. Acta Psychologica, 36, 359-374.

French, J.W. (1965). The relationship of problem-solving styles to the factor composition of tests. Educational and Psychological Measurement, 25, 9-28.

Gagne, R.M. (1962). The acquisition of knowledge. Psychological Review, 69, 355-365.

Gentner, D., and Gentner, D.R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner and A.L. Stevens (Eds.), Mental Models. Hillsdale, NJ: Erlbaum.

Just, M.A., and Carpenter, P.A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. Psychological Review, 92, 137-172.

Kyllonen, P.C., Lohman, D.F., and Snow, R.E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. Journal of Educational Psychology, 76, 130-145.

de Leeuw, J., and Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. Journal of Educational Statistics, 11, 183-196.

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.

Macrae, K.S. (n.d.). Strategies underlying psychometric test responses in young and middle-aged adults of varying educational background. La Trobe University, Australia.

Messick, S. (1984). The psychology of educational measurement. Journal of Educational Measurement, 21, 215-237.

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

Paulson, J. (1985). Latent class representation of systematic patterns in test responses. ONR Technical Report.

Portland: Portland State University. Pelligrino, J.W., Mumaw, R.J., and Shute, V.J. (1985) Analysis of spatial aptitude and expertise. In S.E. Embretson (Ed.), Test Design: Developments in Psychology and Psychometrics. Orlando: Academic Press.

Piaget, J. (1960). The general problems of the psychological development of the child. In J.M. Tanner and B. Inhelder (Eds.), Discussions on Child Developement: Vol. 4. The fourth meeting of the World Health Organization Study Group on the Psychological Developement of the child, Geneva, 1956.

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

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

Schieblechner, H. (1972). Das lernen und losen komplexer denkaufgaben. Zeitschrift fur experimentelle und Angewandte Psychologie, 19, 476-506.

Siegler, R.S. (1981). Developmental sequences within and between concepts. Monograph of the Society for Research in Child Development, Serial No. 189, 46(2).

Tapley, S.M., and Bryden, M.P. (1977). An investigation of sex differences in spatial ability: Mental rotation of three-dimensional objects. Canadian Journal of Psychology, 31, 122-130.

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

Wilson, M.R. (1984). A Psychometric Model of Hierarchical Development. Doctoral dissertation, University of Chicago.

Wilson, M.R. (1985). Measuring Stages of Growth: A Psychometric Model of Hierarchical Development. Occasional Paper No. 19. Hawthorne, Austrailia: Australian Council for Educational Research.

Yamamoto, K. (1987). A hybrid model for item responses. Doctoral dissertation, University of Illinois.

Table 1

Item Features

Item	rotational displacement	salient features
1	60 degrees	3
2	120 degrees	2
3	180 degrees	1

Table 2

Item Difficulty Parameters

Item	Strategy 1	Strategy 2
1	-1.0	2.0
2	0.0	0.5
3	1.0	-1.0

Table 3

Observed and Fitted Response Pattern Counts for Example 2

x	observed frequencies	expected frequencies (Rasch model only)	expected frequencies (2-class model)
0000	143	143.00	143.08
0001	94	98.66	93.95
0010	83	87.12	83.11
0011	101	90.55	101.09
0100	73	72.75	72.78
0101	78	76.62	77.75
0110	65	66.77	65.26
0111	106	93.20	105.98
1000	64	55.46	63.91
1001	54	57.65	54.16
1010	47	50.91	46.75
1011	71	71.06	70.94
1100	39	42.51	39.30
1101	54	59.34	54.07
1110	45	52.40	44.80
1111	83	83.00	83.07

Table 4

Response Pattern Likelihoods and Posterior Probabilities

x	$L(x\|\phi_2)$	$L(x\|\theta_2,\phi_1)$	$L(x\|\theta_2,\phi_1)$	$P(\phi_2\|x)$	$P(\theta_2\|x,\phi_1)$	$P(\theta_3\|x,\phi_1)$
0000	.388	.150	.027	.534	.719	.281
0001	.063	.131	.058	.131	.513	.487
0010	.085	.107	.047	.200	.513	.487
0011	.014	.093	.100	.027	.303	.697
0100	.116	.080	.036	.313	.513	.487
0101	.019	.070	.076	.047	.303	.697
0110	.025	.057	.062	.076	.303	.697
0111	.004	.050	.131	.008	.151	.849
1000	.156	.053	.024	.481	.513	.487
1001	.025	.047	.050	.092	.303	.697
1010	.034	.038	.041	.143	.303	.697
1011	.005	.033	.087	.015	.151	.849
1100	.047	.029	.031	.234	.303	.697
1101	.008	.025	.065	.027	.151	.849
1110	.010	.020	.053	.045	.151	.849
1111	.002	.018	.113	.004	.068	.932

Note: ϕ_1 denotes membership in the class of Rasch responders;
ϕ_2 denotes membership in the class of random guessers;
θ_2 denotes membership in the class of Rasch responders, with θ=-.534;
θ_3 denotes membership in the class of Rasch responders, with θ=.354.

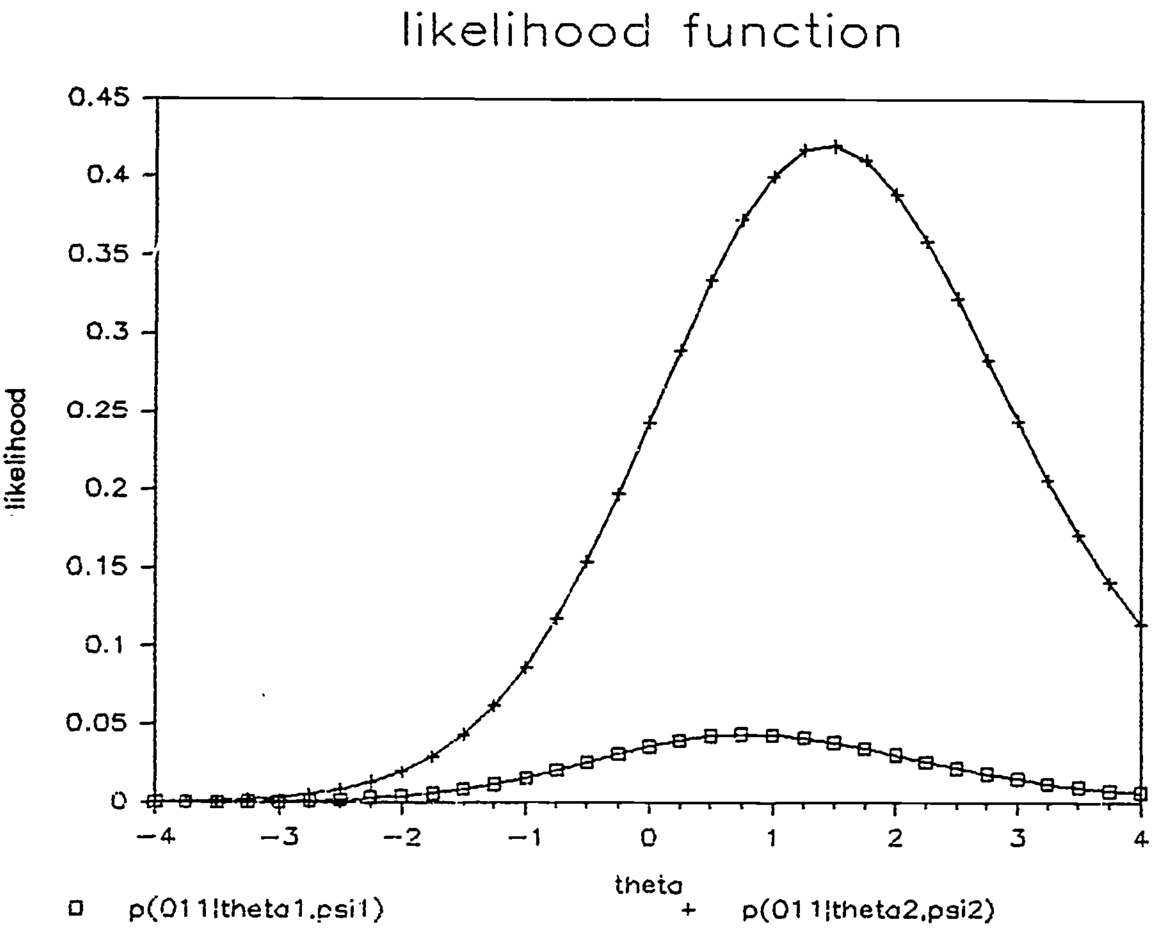
likelihood function
likelihood
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-4
-3
-2
-1
0
1
2
3
4
theta
p(011|theta1,psi1)
p(011|theta2,psi2)

Figure 1

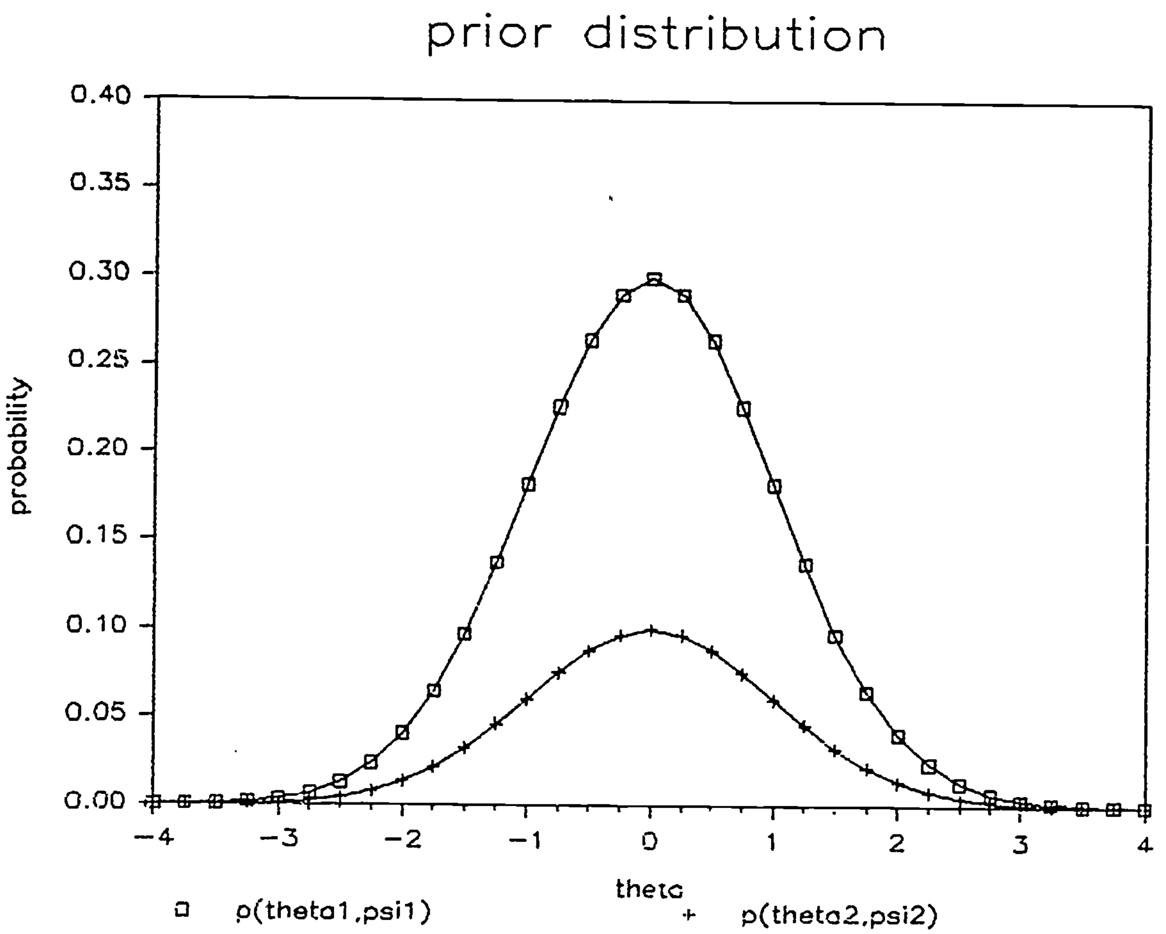
prior distribution
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
probability
−4
−3
−2
−1
0
1
2
3
4
theta
p(theta1,psi1)
p(theta2,psi2)

Figure 2

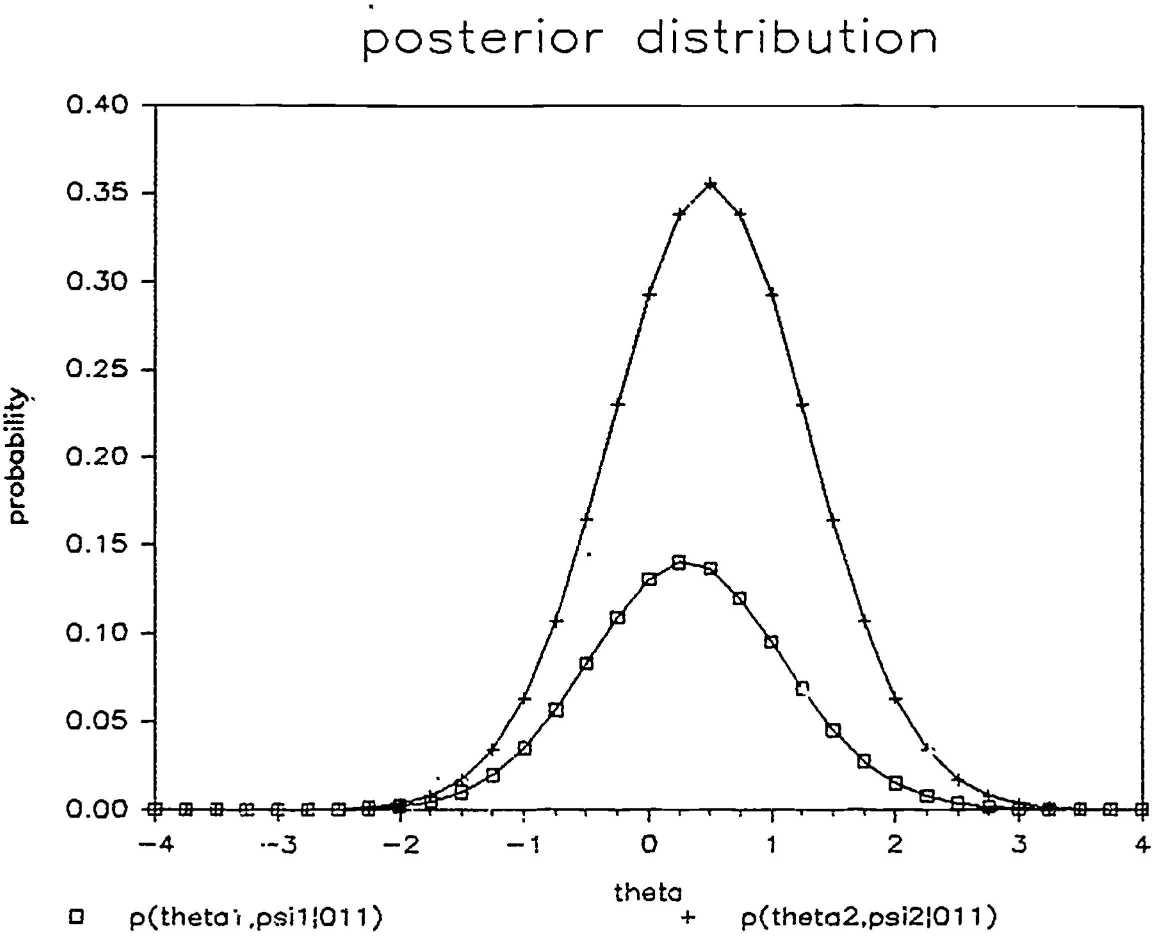
posterior distribution
probability
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-4
-3
-2
-1
0
1
2
3
4
theta
p(theta1,psi1|011)
p(theta2,psi2|011)

Figure 3

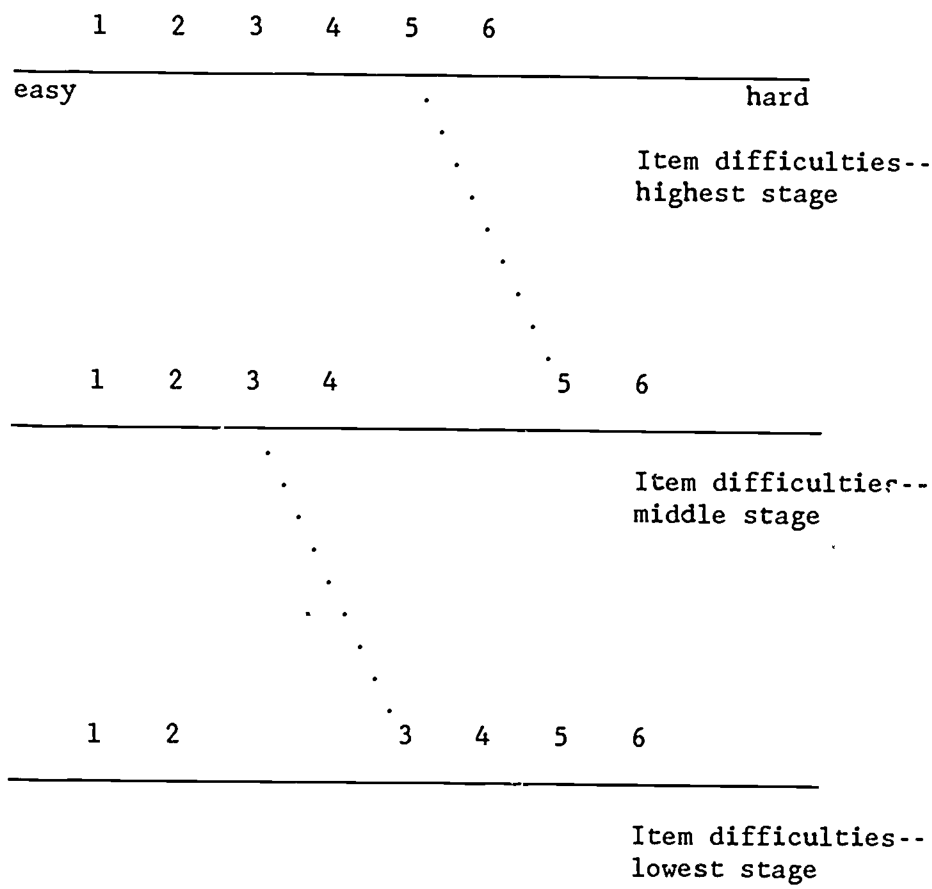

Figure 4

Saltus example: 3 stages, common offset

Dr. Terry Ackerman
American College Testing Programs
P.O. Box 168
Iowa City, IA 52243

Dr. Robert Ahlers
Code N711
Human Factors Laboratory
Naval Training Systems Center
Orlando, FL 32813

Dr. James Algina
University of Florida
Gainesville, FL 32605

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

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 6[illegible]
ISRAEL

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

Dr. Bruce Bloxom
Defense Manpower Data Center
550 Camino El Estero,
Suite 200
Monterey, CA 93943-3231

Dr. R. Darrell Bock
University of Chicago
NORC
6030 South Ellis
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
Naval Training Systems Center
Orlando, FL 32813

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

Dr. Lyle D. Broemeling
ONR Code 1111SP
800 North Quincy Street
Arlington, VA 22217

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
OP 01B7
Washington, DC 20370

Mr. Raymond E. Christal
AFHRL/MOE
Brooks AFB, TX 78235

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

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

Mr. Timothy Davey
University of Illinois
Educational Psychology
Urbana, IL 61801

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Building
University of Maryland
College Park, MD 20742

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

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1k226
Piscataway, NJ 08854

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
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
426 Fraser
Lawrence, KS 66045

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

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

Dr. Pat Federico
Code 511
NPRDC
San Diego, CA 92152-6800

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

Dr. Myron Fischl
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

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 92152-6800

Dr. Alfred R. Fregly
AFOSR/NL
Bolling AFB, DC 20332

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor 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. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Dipl. Pad. Michael W. Habon
Universitat Dusseldorf
Erziehungswissenschaftliches
Universitatsstr. 1
D-4000 Dusseldorf 1
WEST GERMANY

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

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

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

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

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

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

Mr. Dick Hoshaw
OP-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. Dennis E. Jennings
Department of Statistics
University of Illinois
1409 West Green Street
Urbana, IL 61801

Dr. Douglas H. Jones
Thatcher Jones Associates
P.O. Box 6640
10 Trafalgar Court
Lawrenceville, NJ 08648

Dr. Milton S. Katz
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

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

Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Daryll Lang
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Jerry Lehnus
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

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

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541

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. Milton Maier
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

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

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

Dr. Robert McKinley
Educational Testing Service
20-P
Princeton, NJ 08541

Dr. James McMichael
Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

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

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

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

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

Deputy Technical Director
NPRDC Code 01A
San Diego, CA ...152-6800

Director Training Laboratory,
NPRDC (Code 05)
San Diego, CA 92152-6800

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152-6800

Director, Human Factors
& Organizational Systems Lab,
NPRDC (Code 07)
San Diego, CA 92152-6800

Fleet Support Office,
NPRDC (Code 301)
San Diego, CA 92152-6800

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

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

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern California
Los Angeles, CA 90089-0031

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

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

Office of Naval Research,
Code 125
800 N. Quincy Street
Arlington, VA 22217-5000

Assistant for MPT Research,
Development and Studies
OP 01B7
Washington, DC 20370

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

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Randolph Park
Army Research Institute
5001 Eisenhower Blvd.
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

Administrative Sciences Department,
Naval Postgraduate School
Monterey, CA 93940

Department of Operations Research,
Naval Postgraduate School
Monterey, CA 93940

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

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

Dr. Barry Riegelhaupt
HumRRO
1100 South Washington Street
Alexandria, VA 22314

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
3108 AustinPeay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Dan Segall
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 E. Snow
Department of Psychology
Stanford University
Stanford, CA 94306

Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

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

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

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 Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

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 92152-6800

Dr. John Tangney
AFOSR/NL
Bolling AFB, DC 20332

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
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

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 92152-6800

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

Dr. 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 92152-6800

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 12
Navy Personnel R&D Center
San Diego, CA 92152-6800

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
NRC GF-176
2101 Constitution Ave
Washington, DC 20418

Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152-6800

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

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

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
Computer-based Education
Research Laboratory
University of Illinois
Urbana, IL 61801

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

Dr. Joseph L. Young
Memory & Cognitive
Processes
National Science Foundation
Washington, DC 20550

Dr. Anthony R. Zara
National Council of State
Boards of Nursing, Inc.
625 North Michigan Ave.
Suite 1544
Chicago, IL 60611