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)
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

Reproduction in whole or in part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.
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.
Modeling Item Responses When Different Subjects Employ Different Solution Strategies

Robert J. Mislevy
Educational Testing Service
and
Norman Verhelst
CITO
(National Institute for Educational Measurement)
Arnhem, The Netherlands
October 1987

1The 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
Different Strategies

(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

\[
p(x_{ij} | \theta_i, \beta_j) = \frac{f(\theta_i, \beta_j)}{f(\theta_i, \beta_j) + 1 - f(\theta_i, \beta_j)}
\]

where \( \theta_i \) and \( \beta_j \) 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 \( x_i = (x_{i1}, \ldots, x_{in}) \) of person \( i \) to \( n \) items is the product of \( n \) expressions like (1):
Different Strategies

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

It may be 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

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

and the model for item parameters is linear:

\[
\beta_j(\alpha) = \sum_{m=1}^{M} Q_{jm} \alpha_m = Q'_j \alpha.
\]

The elements of \( \alpha \) 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 $\phi_i$ 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 $x_i$ as

$$p(x_i | \phi_i, \theta_i, \alpha) = \prod_{k=1}^{K} [\pi_k(\theta_{ik}, \beta_{jk})]^{x_{ij}}[1 - \pi_k(\theta_{ik}, \beta_{jk})]^{1-x_{ij}} \phi_{1k}$$

(2)

where $\beta_{jk} = \beta_{jk}(\alpha)$ 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 $Q_{jk}$, $k=1, \ldots, K$, 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 $\alpha_k$, the basic parameter vector associated with Strategy $k$; the weights $Q_{jkm}$ indicate the degree to which each of the features $m$, as relevant 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, $\phi_i = (\phi_{i1}, \phi_{i2})$, $\phi_{ik}$ takes the value 1 if Subject $i$ employs
Different Strategies

Strategy $k$ and $0$ if not. In the second, $\theta_i = (\theta_{i1}, \theta_{i2})$, $\theta_{i k}$ characterizes the proficiency of Subject $i$ if he employs Strategy $k$. Only one of the elements of $\theta_i$ 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:

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

Here $\theta_{i1}$ is the proficiency of Subject $i$ at solving tasks by means of the rotational strategy, and $\beta_{j1}$ 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 $\alpha_{11}$ is the incremental increase in difficulty for each increment in rotation; and $\alpha_{12}$ 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
Different Strategies

\( \beta_{j1} \) that are in effect under Strategy 1 are as shown in the second column of Table 2.

\[ \begin{align*}
\beta_{j2} &= Q_{j21} \alpha_{21} + \alpha_{22},
\end{align*} \]

where \( Q_{j21} \) is the number of salient features, \( \alpha_{21} \) is the incremental contribution to item difficulty of an additional feature, \( \alpha_{22} \) is a constant term, and \( Q_{j22} = 1 \) implicitly for all items. If \( \alpha_{21} = 1.5 \) and \( \alpha_{22} = -2.5 \), 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...
Different Strategies

11

Strategy 2 but easy under Strategy 1. Figure 1 plots the likelihood function for the response vector 011 under both strategies; that is, \( p(x=(011)|\theta_1, \phi_k=1) \) for \( k=1,2 \) as a function of \( \theta_1 \) and \( \theta_2 \) respectively. The maximum of the likelihood under Strategy 2 is about eight times as high as the maximum attained under Strategy 1.

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) \( \theta_1 \) and \( \theta_2 \) 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.

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

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

\[ p[x|\phi_k=1, \theta, \beta_k(\alpha)] = \prod_{j} \exp\left(\xi_{ij} [\theta - \beta_{jk}(\alpha)]/(1 + [\theta - \beta_{jk}(\alpha)]) \right). \]

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 \( \theta_k \) 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 \( \theta_1 \) given \( \phi_1=1 \) has a mean and standard deviation of about .32 and .80. Corresponding values for the distribution of \( \theta_2 \) given \( \phi_2=1 \) are .50 and .81.

Parameter Estimation

This section discusses estimation procedures for mixtures of IRT models. A two-stage procedure is described. The first stage
integrates over $\theta$ and $\phi$ 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 $\theta$ 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 $p(\theta,|\phi_k=1,x)$.

Stage 1: Estimates of Structural Parameters

Equation 2 gives the conditional probability of the response vector $x$ given $\theta$ and $\phi$, or $p(x|\theta,\phi,\alpha)$. Consider a population in which strategies are employed in proportions $\pi_k$ 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

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

For brevity, let $\xi$ denote the extended vector of all structural parameters, namely $(\alpha,\pi,\eta)$. The loglikelihood for $\xi$ 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:

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

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

$$+ \sum_i \sum_k \phi_{ik} \log \pi_k .$$

Let $S$ be the vector of first derivatives, and $H$ the matrix of second derivatives, of $\lambda$ with respect to $\xi$. 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 $\xi^0$ by adding the correction term $-H^{-1}S \bigg|_{\xi=\xi^0}$. 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.
Different Strategies

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 $g_k$, 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, $g_k(\theta_{k1}|\eta_k)$. Define the subject variable $\psi_i = (\psi_{i1l}, \ldots, \psi_{iKL(k)})$, a vector of length $\Sigma_k L(k)$ where the element $\psi_{ikl}$ is 1 if the proficiency of Subject i under Strategy k is $\theta_{kl}$ and 0 if not. There are a total of K 1s in $\psi_i$, 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

$$\lambda = \sum_{i} \sum_{k} \phi_{ik} \sum_{l} \psi_{ikl} \log p[x_i | \theta_{k1}, \ldots, \theta_{kL}, \phi_{k1L}, \phi_{kL-1}, \phi_{kL}(\alpha)]$$

$$+ \sum_{i} \sum_{k} \phi_{ik} \sum_{l} \psi_{ikl} \sum_{k} g_k(\theta_{kl}|\eta_k)$$

$$+ \sum_{i} \sum_{k} \phi_{ik} \log \pi_k \quad \text{(6)}$$
If values of $\phi$ and $\psi$ were observed along with values of $x$, ML estimation of $\xi$ from (6) would be simpler. The basic parameter $\alpha$ appears only in the first term on the right side of (6), so that maximizing with respect to $\alpha$ need address only that term. When $\alpha$ consists of distinct subvectors for each strategy, even these subvectors lead to distinct maximization problems of lower order.

The subpopulation parameters $\eta$ appear in only the second term, separating them in ML estimation; they too lead to even smaller separate subproblems if $\eta$ consists of distinct subvectors for each strategy. The population proportions $\pi$ appear in only the last term. Unless they are further constrained, their ML estimates are simply observed proportions. The values of $\theta$ 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 $\alpha$ and $\eta$ 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 $\phi_i$ and $\psi_i$ are replaced by their conditional expectations given $x_i$ and provisional estimates of $\xi$, say $\xi^0$. The
Different Strategies

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

$$l_{ikl} = p[x_i|\theta_k, \phi_k, \beta_k(a)] g_k(\theta_k, \eta_k) \pi_k .$$

The required conditional expectations are obtained as

$$\psi_{ikl}^0 = E(\psi_{ikl}|x_i, \xi-\xi^0) = l_{ikl}^0 / \sum_{\ell'} l_{ikl'}^0$$

(7)

and

$$\phi_{ik}^0 = E(\phi_{ik}|x_i, \xi-\xi^0) = l_{ikl}^0 / \sum_{k'\ell'} l_{ik'\ell'}^0$$

(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' \( \theta \) and \( \phi \) 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 \( \xi \) are accurately estimated, the posterior density of the parameters of examinee \( i \) is approximately

\[
p(\theta_{ik} = \hat{\theta}, \phi_{ik} = 1 | x_i, \hat{\xi}) \propto p(x_i | \phi_{ik} = 1, \theta_{ik}, \beta_k(\alpha)) \pi_k g_k(\theta | \eta_k),
\]

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

\[
P(\phi_{ik} = 1 | x_i, \hat{\xi}) = \int p(\theta_{ik} = \hat{\theta}, \phi_{ik} = 1 | x_i, \hat{\xi}) \, d\theta.
\]
The examinee's posterior mean and variance for a given strategy class, given that that was the strategy employed, are approximated by

$$\bar{\theta}_{ik} = \int \theta p(\theta_{ik} = \theta, x_{ik} = 1 | x_i, \xi) \, d\theta / p(\phi_{ik} = 1 | x_i, \xi)$$

and

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

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

$$P(x_1|\xi) = \sum_{k=1}^{2} P(x_1|\phi_{k-1},\xi) \pi_k,$$

where Strategy Class 1 is the Rasch model and Class 2 is random guessing. The composition of $\xi$ is now described. It includes first the strategy proportions $\pi_1$ and $\pi_2$. For the Rasch class, the basic parameters $\alpha_1$ 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 $\alpha_2$ 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

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

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 $n_1 = .833$ and $n_2 = .167$. The Rasch item parameters were $\alpha_1 = (b_1, \ldots, b_4) = (-.511, -.105, .182, .405)$. 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:

<table>
<thead>
<tr>
<th>Point</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.204</td>
<td>.08</td>
</tr>
<tr>
<td>-.357</td>
<td>.17</td>
</tr>
<tr>
<td>.095</td>
<td>.25</td>
</tr>
<tr>
<td>.262</td>
<td>.25</td>
</tr>
<tr>
<td>.470</td>
<td>.17</td>
</tr>
<tr>
<td>.642</td>
<td>.08</td>
</tr>
</tbody>
</table>

The rates of correct response for the random guessers on the four items were $\alpha_2 = (c_1, \ldots, c_4) = (.30, .35, .20, .15)$. 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.
Different Strategies

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
Different Strategies

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: $\theta_2$ and $\theta_3$.
- 2 free values for weights at the three points in the Rasch
distribution: $\omega_1$, $\omega_2$, and $\omega_3$, where $\Sigma \omega_k = 1$.
- 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: $\pi_2$.

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
$\pi_2$, starting values of .10, .15, and .20 were used in three
different runs. All runs converged to the same solution:
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.
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 \( p(θ|x,φ_{1-1}) \), 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
Different Strategies

26

class can be inferred from Table 4 as $1 - P(\phi_2 - 1|\pi)$. 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's (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
Different Strategies

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 test 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
elements, 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, \( \phi \) 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 $\alpha$ 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
Different Strategies

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 strategies. 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
Different Strategies

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


Gentner, D., and Gentner, D.R. (1983). Flowing waters or teeming


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


Different Strategies


Different Strategies

36


Different Strategies

Table 1
Item Features

<table>
<thead>
<tr>
<th>Item</th>
<th>Rotational Displacement</th>
<th>Salient Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>60 degrees</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>120 degrees</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>180 degrees</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2
Item Difficulty Parameters

<table>
<thead>
<tr>
<th>Item</th>
<th>Strategy 1</th>
<th>Strategy 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>1.0</td>
<td>-1.0</td>
</tr>
</tbody>
</table>
### Different Strategies

#### Table 3

Observed and Fitted Response Pattern Counts for Example 2

<table>
<thead>
<tr>
<th>x</th>
<th>observed frequencies</th>
<th>expected frequencies (Rasch model only)</th>
<th>expected frequencies (2-class model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>143</td>
<td>143.00</td>
<td>143.08</td>
</tr>
<tr>
<td>0001</td>
<td>94</td>
<td>98.66</td>
<td>93.95</td>
</tr>
<tr>
<td>0010</td>
<td>83</td>
<td>87.12</td>
<td>83.11</td>
</tr>
<tr>
<td>0011</td>
<td>101</td>
<td>90.55</td>
<td>101.09</td>
</tr>
<tr>
<td>0100</td>
<td>73</td>
<td>72.75</td>
<td>72.78</td>
</tr>
<tr>
<td>0101</td>
<td>78</td>
<td>76.62</td>
<td>77.75</td>
</tr>
<tr>
<td>0110</td>
<td>65</td>
<td>66.77</td>
<td>65.26</td>
</tr>
<tr>
<td>0111</td>
<td>106</td>
<td>93.20</td>
<td>105.98</td>
</tr>
<tr>
<td>1000</td>
<td>64</td>
<td>55.46</td>
<td>63.91</td>
</tr>
<tr>
<td>1001</td>
<td>54</td>
<td>57.65</td>
<td>54.16</td>
</tr>
<tr>
<td>1010</td>
<td>47</td>
<td>50.91</td>
<td>46.75</td>
</tr>
<tr>
<td>1011</td>
<td>71</td>
<td>71.06</td>
<td>70.94</td>
</tr>
<tr>
<td>1100</td>
<td>39</td>
<td>42.51</td>
<td>39.30</td>
</tr>
<tr>
<td>1101</td>
<td>54</td>
<td>59.34</td>
<td>54.07</td>
</tr>
<tr>
<td>1110</td>
<td>45</td>
<td>52.40</td>
<td>44.80</td>
</tr>
<tr>
<td>1111</td>
<td>83</td>
<td>83.00</td>
<td>83.07</td>
</tr>
</tbody>
</table>
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| 0.388         | 0.150                    | 0.027                    | 0.534         | 0.719                    | 0.281                    |
| 0001| 0.063         | 0.131                    | 0.058                    | 0.132         | 0.513                    | 0.487                    |
| 0010| 0.085         | 0.107                    | 0.047                    | 0.200         | 0.513                    | 0.487                    |
| 0011| 0.014         | 0.093                    | 0.100                    | 0.027         | 0.303                    | 0.697                    |
| 0100| 0.116         | 0.080                    | 0.036                    | 0.313         | 0.513                    | 0.487                    |
| 0101| 0.019         | 0.070                    | 0.076                    | 0.047         | 0.303                    | 0.697                    |
| 0110| 0.025         | 0.057                    | 0.062                    | 0.076         | 0.303                    | 0.697                    |
| 0111| 0.004         | 0.050                    | 0.131                    | 0.008         | 0.151                    | 0.849                    |
| 1000| 0.156         | 0.053                    | 0.024                    | 0.481         | 0.513                    | 0.487                    |
| 1001| 0.025         | 0.047                    | 0.050                    | 0.092         | 0.303                    | 0.697                    |
| 1010| 0.034         | 0.038                    | 0.041                    | 0.143         | 0.303                    | 0.697                    |
| 1011| 0.005         | 0.033                    | 0.087                    | 0.015         | 0.151                    | 0.849                    |
| 1100| 0.047         | 0.029                    | 0.031                    | 0.234         | 0.303                    | 0.697                    |
| 1101| 0.008         | 0.025                    | 0.065                    | 0.027         | 0.151                    | 0.849                    |
| 1110| 0.010         | 0.020                    | 0.053                    | 0.045         | 0.151                    | 0.849                    |
| 1111| 0.002         | 0.018                    | 0.113                    | 0.004         | 0.068                    | 0.932                    |

Note: $\phi_1$ denotes membership in the class of Rasch responders; $\phi_2$ denotes membership in the class of random guessers; $\theta_2$ denotes membership in the class of Rasch responders, with $\theta=-.534$; $\theta_3$ denotes membership in the class of Rasch responders, with $\theta=.354$. 

Different Strategies
Different Strategies

likelihood function

\[ \text{likelihood function} \]

\[ p(011 | \theta_1, \psi_1) \]

\[ p(011 | \theta_2, \psi_2) \]

\[ p(011 | \theta_2, \psi_2) \]

Figure 1
Different Strategies

prior distribution

\[ p(\theta_1, \psi_1) + p(\theta_2, \psi_2) \]

Figure 2
Different Strategies

posterior distribution

Figure 3

\[ p(\theta_1, \psi_1 | 011) + p(\theta_2, \psi_2 | 011) \]
Different Strategies

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>easy</td>
<td>hard</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item difficulties--
highest stage

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item difficulties--
middle stage

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Item difficulties--
lowest stage

Figure 4
Saltus example: 3 stages, common offset
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
Universität Düsseldorf
Erziehungswissenschaftliches
Universitätsstr. 1
D-4000 Düsseldorf 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. 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 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 08541

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

Deputy Technical Director
NPRDC Code 01A
San Diego, CA 92152-6800

Director Training Laboratory,
NPRDC (Code 05)
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
AF-HRL/MAAN
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