DOCUMENT RESUME

ED 395 036 TM 025 056

AUTHOR Yamamoto, Kentaro; Everson, Howard T.
TITLE Modeling the Mixture of IRT and Pattern Responses by a Modified HYBRID Model.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-95-16
PUB DATE Jul 95
NOTE 29p.; Paper presented at a symposium titled "Applications of Latent Trait and Latent Class Models in the Social Sciences" (Hamburg, Germany, May 1994).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Ability; College Students; Estimation (Mathematics); Higher Education; *Item Response Theory; *Psychometrics; Reading Comprehension; *Reading Tests; *Timed Tests
IDENTIFIERS *HYBRID Model; Item Parameters; Latent Class Models; *Response Patterns; Speededness (Tests)

ABSTRACT

This study demonstrates the utility of a HYBRID psychometric model, which incorporates both item response theoretic and latent class models, for detecting test speededness. The model isolates where in a sequence of test items examinee response patterns shift from one providing reasonable estimates of ability to those best characterized by a random response pattern. The study applied the HYBRID model to three distinct data sets: (1) simulated data representing the performance of 3,000 examinees on a 70-item test; (2) data from a statewide field test (n=5,997) of a 40-item reading comprehension test with a fixed time limit; and (3) data from a study of urban university students (n=752) who took a similar 40-item reading comprehension test under varying time limits. The HYBRID model successfully identified "switch points" in examinees' item response patterns in all three data sets. This paper discusses applications of the model used to detect speededness and to provide adjusted estimates of item parameters and examinee abilities. (Contains 4 tables, 4 figures, and 38 references.) (Author)

RR-95-16

RESEARCH REPORT

MODELING THE MIXTURE OF IRT AND PATTERN RESPONSES BY A MODIFIED HYBRID MODEL

Kentaro Yamamoto
Howard T. Everson

BEST COPY AVAILABLE

Educational Testing Service
Princeton, New Jersey
July 1995

Modeling the Mixture of IRT and Pattern Responses by a Modified Hybrid Model[1]

Kentaro Yamamoto[2]
Educational Testing Service

Howard T. Everson[3]
The College Board

March 1995

[1] This paper was presented at the IPN symposium entitled: *Applications of Latent Trait and Latent Class Models in the Social Sciences.* at Sankelmark, University of Kiel, Hamburg , May 1994.

[2] Dr. Yamamoto's work was supported by the *Educational Testing Service's* Program Research Planning Council.

[3] Dr. Everson's work was supported, in part, by a Postdoctoral Fellowship from the *Educational Testing Service.*

Modeling the Mixture of IRT and Patterne Responses by a Modified Hybrid Model

Abstract

This study demonstrates the utility of a HYBRID psychometric model, which incorporates both item response theoretic and latent class models, for detecting test speededness. The model isolates where in a sequence of test items examinee response patterns shift from ones providing reasonable estimates of ability to those best characterized by a random response pattern. The study applied the HYBRID model to three distinct data sets: (1) simulated data representing the performance of 3,000 examinees on a 70-item test; (2) data from a statewide field test (n=5997 of a 40 item reading comprehension test with a fixed time limits; and (3) data from a study of urban university students (n=752) who took a similar 40-item reading comprehension test under varying time limits. The HYBRID model successfully identified "switch points" in examinee's item response patterns in all three data sets. This paper discusses applications of the model used to detect speededness and to provide adjusted estimates of item parameters and examinee abilities.

INTRODUCTION

It is no secret that many standardized tests used in educational settings are administered under timed conditions. As a result, it is rare that *all* examinees have *all* the time they need when tested. Critics are quick to point out that arbitrary limits to testing time may be unfair to some groups of examinees, particularly groups for whom English is a second language. Some examinees, it is argued, may simply run out of time because they work slowly, or because they spend a good deal of precious testing time deciphering the test's language. Issues of test bias may arise when examinees from one or another minority group are differentially affected by the test's time limits. Thus, a central problem for test developers and test users--particularly when a test is scored "rights only" (i.e., the total number of correct responses) and examinees are not discouraged from responding randomly at the end of a test (i.e., guessing)--is determining the appropriate time limit for a test.

Unfortunately, model-based methods for determining appropriate time-limits during test development have remained elusive, particularly when examinees are not discouraged from guessing and the proportion of examinees not reaching the last few items is low. For example, the majority of approaches available in the psychometric literature are analyses of patterns of "not reached" items. When guessing occurs, surface examinations of the examinee response patterns, i.e., the number and position of items attempted, reveal little about the solution strategies used by examinees as testing time is exhausted. Do all examinees have sufficient time to attempt all the items on the test, or do some respond randomly as time runs out? From the test development perspective, what is needed is a psychometric model that serves to define and measure the degree of speededness when tests are scored "rights only".

The purpose of this study was to demonstrate the utility of the HYBRID model (Yamamoto, 1989)--a psychometric model that combines *both* item response theoretic and latent class approaches--for detecting speededness in such a test. The model's utility, we argue, derives from its ability to detect the point in a test where a significant proportion of examinees "switch" response strategies from meaningful (ability driven) responses to random respondes. By applying the model to both simulated and field test data, we show how the HYBRID model can be used in the test development phase to determine the optimal length or time-limit of a test, thereby strengthening the measurement properties of the test.

In the following section, we briefly review a number of earlier attempts to develop model-based approaches for estimating speededness, and discuss their limitations. We then describe the HYBRID model and show how it was extended to help identify the "switch points" (i.e., points in a test where examinees response patterns "switch" from meaningful to random respondes) presumably related to the tests length and/or time limits. This description is followed by an overview of the research design, including descriptions of the simulated and field test data analyzed in this study. We conclude by summarizing the results of our analyses and discussing the utility of the HYBRID model for estimating the effects of test length and test speededness during the test development phase.

Earlier Approaches

More than forty years ago, Cronbach and Warrington (1951) argued that ..."test theory will be clarified if we can determine and measure degree of speeding" (p.184). Over the years a number of researchers have attempted to develop rules of thumb and indices for measuring test speededness. The conventional measure of test speededness is the proportion of examinees attempting all test items within the allotted time period. According to criteria used by commercial test developers, including ETS, (Nunnally, 1978; Swineford, 1974) a test may be considered unspeeded if (1) virtually all examinees reach 75% of the items, and (2) at least 90% of the candidates respond to the last item. In a discussion of time limits on standardized tests, Nunnally (1978) speaks of a "comfortable time limit," which he defines as "the amount of time required for 90% of the persons to complete a test under power conditions" (1978, p.632). In general, the number of items completed by 80% of the test takers has served as an index of the relative speededness of a test (Schmitt et al., 1991). This index of speededness, however, is much less useful for "rights only" scored tests, where examinees are encouraged by virtue of the scoring protocol to guess randomly as testing time runs out. Ignoring the random responses of "test-wise" examinees, no doubt, underestimates test speededness.

Modeling only omitted responses may also provide biased ability estimates. Ability estimates derived from item response theory (IRT; Lord, 1980), for example, do not explicitly incorporate speededness estimates. According to the conventional argument, this implies that the IRT based ability estimate is unaffected by variations in speededness or by testing time limits, although it is commonly assumed that performance levels decline when testing time is insufficient.

A number of earlier studies (Bejar, 1985; Donlon, 1980; Secolsky, 1989; Secolsky and Steffen, 1990) bear directly on this issue. Bejar (1985), working within an IRT framework, assumed that when speed was a factor, the less able examinees would leave the more difficult test items for last and, therefore, could be characterized by a random response pattern. He proposed an index of speededness that compared the observed performance on the most difficult test items with the performance predicted by an overall IRT model. Bejar's index, analogous to a chi-square statistic, was calculated for several ability levels. As Bejar himself noted, the method was circular because the IRT parameters were estimated using the items that were assumed to be affected by speededness. More important, the speededness index did not reflect the fit of the three-parameter IRT model. He reported, for example, model misfit in the extreme high and low ability regions, regions where the IRT model parameters were least accurately estimated. As a result, Bejar's index failed to detect speededness in the more important regions of the ability distribution.

Secolsky (1989) attempted to address the issue of test speededness by using techniques based on regression methods. Secolsky worked from the premise that under power (non-speeded) test conditions, scores on items early in the test would correlate highly with scores on items at the end of the test (i.e., scores on beginning test items should predict scores on test items at the end). Under speeded conditions, the expected relationship between the two portions of the test would be different.

He examined data from four administrations of the Test of English as a Foreign Language (TOEFL) and concluded that the test was slightly speeded because the observed scores of a subset of examinees on the last 4 to 6 test items were significantly lower than the scores predicted by performance on the first 4 to 6 items. Unfortunately, regression methods based on such a small number of items are less than reliable, and may be subject to errors of classification based on this uncertainty. Conclusions about test speeededness may be an artifact of the unreliability inherent in this application of regression methods. In addition, the techniques used by Secolsky (1989) were somewhat arbitrary in that they examined speededness at a single point in the test sequence, the last 4 to 6 items, and ignored individual differences in speed of responding.

In sum, these earlier approaches shared a number of shortcomings: (1) they did not examine the performance of the procedures when the test was unspeeded; (2) they examined speededness at somewhat arbitrary points in the test; (3) they were, for the most part, unconcerned with the bias in the model parameters that the test's speededness may have introduced; and (4) these approaches did not address the detection of differential speededness by subpopulations of examinees. Recent developments in IRT modeling, i.e., the development of a "hybrid" model (see Yamamoto, 1990 for details) address these problems by changing the essential question from "Is a test speeded?" to "How speeded is a test?" In the next section, we discuss extensions to the HYBRID model that permit estimates of the proportion of examinees who shift or switch to a random response pattern at various points in the test item sequence. This model-based approach also enables us to examine carefully the effects of speededness on both the item and ability parameter estimates.

THE HYBRID MODEL

Both classical test theory and item response theory (IRT), including the one-, two-, and three-parameter normal and logistic models, use a single measurement model to characterize the examinee responses. These traditional psychometric models are efficient methods for ordering examinees on a unidimensional ability continuum, and they work reasonably well for tests where examinees use essentially the same strategy to solve the items. They are less well suited to testing situations that are decidedly multidimensional, or when examinees switch solution strategies at various points in the test (Kyllonen, Lohman, and Snow, 1984; Secolsky, 1989). Currently, there are two psychometric models that incorporate multiple response strategies--the HYBRID model (Yamamoto, 1989) and the Mixed Strategies model (Mislevy and Verhelst, 1990). When tests are used for diagnostic purposes or academic placement decisions, more information about examinees' cognitive processing characteristics is often required Thus, there is a need for psychometric models that capture the qualitative nature of an individual's performance on a test.

The HYBRID model supplements the discrete latent class model (LCM) of item responses with a continuous IRT model. In its basic conceptualization, the model presupposes a class of examinees whose responses are characterized well by an IRT model. For such a group of examinees, differences in their response patterns do not typically indicate qualitative differences in their solution strategies. The response pattern of other examinees, however, may be more accurately described by a model that places test takers in discrete classes that are associated with qualitatively different solution strategies. The HYBRID model uses both psychometric approaches to achieve an optimal fit for a sample of examinees. For detecting test speededness, the model can be extended to capture strategy switching; in this way a subset of examinees' responses is best described by a latent class model (i.e., a guessing class) while the remainder of the responses fit an IRT model.

The HYBRID model assumes that conditional independence holds for both the IRT and the latent class groups. The model produces three sets of parameters: (1) IRT parameters--i.e., a set of item parameters for each item and an ability parameter for each examinee; (2) an estimate of the proportion of the population of examinees in the IRT and latent classes; and (3) a set of conditional probability estimates for each latent class. The ability parameter, however, is useful only for the proportion of examinees that are characterized adequately by the IRT model. Parameter estimation is done via the Marginal Maximum Likelihood (MML) method (Bock & Aitkin, 1981; Mislevy, 1983). Additional descriptions of the MML method are found in Harwell, Baker, and Zwarts (1988) and Yamamoto (1989).

The HYBRID model uses conventional IRT parameters--i.e., a one-,two-, or three-parameter logistic. For example, the conditional probability of a correct response to item i under the two-parameter logistic IRT model is $P(x_i=1 \mid \theta_j,\beta_i)$,with parameter values $\beta_i=(a_i,b_i)$, and examinee ability θ_j. This is given by

$$P(x_i = 1|\theta_j,\beta_i) = \frac{1}{1.0 + \exp(-Da_i(\theta_j - b_i))} \quad (1)$$

The probability of a correct response to item i for an examinee in a latent class k (k=2,.....K, if there are K-1 latent classes) is denoted by $P(x_i=1 \mid \gamma=k)$. A large number of latent classes have been proposed in the past. They differ in the type of constraints used for conditional probabilities. Some latent classes can be a function of P(q),Yamamoto (1987) proposed several such constraints in addition to more traditional constraints. Some of them can be expressed as $p(x_i=1|\gamma=\kappa,\theta)$. In this paper, we limit our discussion to the standard latent class models.

By the assumption of conditional independence in the IRT model as well as in the latent class groups, the conditional probabilities of observing a response vector $\underline{\mathbf{x}}$ under the IRT class and the K-1 latent classes are:

$$P(\underline{x} \mid \theta,\beta) = \prod_{i=1}^{I} P(x_i = 1|\theta,\beta_i)^{x_i}(1 - P(x_i = 1|\theta,\beta_i))^{1-x_i} \qquad (2)(IRT)$$

$$P(\underline{x} \mid \gamma = k) = \prod_{i=1}^{I} P(x_i = 1|\gamma = k)^{x_i}(1 - P(x_i = 1|\gamma = k))^{1-x_i} \quad (3)\ (LC)$$

Let γ=1 indicate the class modeled by IRT, and let γ=2,....K indicate the K-1 latent classes. Further, let P(γ=k) denote the probability of membership in class k. The marginal probability of observing a response pattern, $\underline{\mathbf{x}}$, given the model parameters β and Γ, is the summation of conditional probabilities over all classes including the IRT group and the latent classes, and is expressed by

$$P(\underline{x} \mid \beta) = \sum_{k=1}^{K} P(\underline{x} \mid \beta,\gamma = k)P(\gamma = k) \qquad (4)$$

$$= \int_{\theta} P(\underline{x} \mid \theta,\beta)f(\theta)d\theta P(\gamma = 1) + \sum_{k=2}^{K} P(\underline{x} \mid \beta,\gamma = k)P(\gamma = k). \qquad (5)$$

Since calculating the integral in the above function and succeeding derivations is cumbersome, a method based on Dempster, Laird, and Rubin's (1977) EM algorithm is used for actual parameter estimation.

Extending the Model

When a test is speeded and scoring is based on the number of correct responses, patterned responses are often observed at the end of the test. This occurs, for example, when an examinee runs out of time in a multiple-choice type test option A may be selected for the last n items. Yet unless the algorithm of patterned responses is obvious, it is often quite difficult to determine whether a portion of the overall responses is patterned or not.

$$P(x_{ij} = 1|\theta,\beta_i,k) = (1 + \exp(-Da_i(\theta - b_i)))^m C_i^{m+1} \qquad (6)$$

Where m=-1 when $i<k_j$, and m=0 when $i=k_j$, x_i is a dichotomous response (0=wrong, 1=right) on item i by examinee j, β_i is a vector of item parameters (a_i,b_i), θ_j is the ability of examinee j, and k_j indicates the last item answered by examinee j under the IRT model. The likelihood of observing $\underline{x}_j$ given θ_j and k_j is

$$P(\mathbf{x}_j = 1|\theta,\beta_i,k) = \prod_{i=1}^{k_j} P(\theta_j,\beta_i)^{x_{ij}} Q(\theta_j,\beta_i)^{1-x_{ij}} \prod_{i=k_j+1}^{I} c_i^{x_{ij}}(1-c_i)^{1-x_{ij}} \qquad (7)$$

(Notice that for those examinees who did not switch response strategy, the likelihood is identical to those of the IRT only model given in Equation 6.)

Moreover, the marginal probability of observing $\underline{x}_j$ given the item parameters is,

$$P(\mathbf{x}_j|\mathbf{B}) = \sum_K^I \int_\theta P(\mathbf{x}_j|\theta,\mathbf{B},k) f(\theta|k) d\theta f(k). \tag{8}$$

Where $f(\theta \mid k)$ represents the conditional probability of θ given a switch point k, and f(k) is the distribution of switch points in the population.

The joint likelihood of the parameters given the observed response matrix $\underline{X}=(\underline{x}_1,\underline{x}_2,\ldots\underline{x}_j)$ from a total of J examinees is

$$L(\mathbf{B} \mid \mathrm{X}) = \prod_j^J P(\mathbf{x}_j \mid \mathbf{B}). \tag{9}$$

The IRT item parameters can be estimated to maximize the marginalized likelihood function in Equation 9 using an iterative method such as the Newton-Raphson (N-R: Kendall and Stuart, 1967) method. The N-R method can be described as $\mathbf{P}^{n+1} = \mathbf{P}^n - \mathbf{D}_2^{-1} * \mathbf{D}_2$, where $\mathbf{P}^{n+1}$ is a vector of updated parameters from $\mathbf{P}^n$ by an amount designated by the function of $\mathbf{D}_2$ (a matrix of second derivatives) and $\mathbf{D}_1$ (a vector of first derivatives). However, $\mathbf{D}_2$ can be quite large, and the off diagonals need not be zero. Consequently, standard application of the N-R method would present too great a computational burden. Bock and Aitkin (1981) advanced the idea of using the EM algorithm (Dempster, Laird, and Rubin, 1977) with probit analysis inner cycles in the area of IRT parameter estimation by replacing continuous θ with discrete θ points, chosen as convenient quadrature points for the integration. Thus, with respect to u as a model parameter for either an item or a population density, the first derivative of the log-likelihood of Equation 9 can be expressed as

$$\frac{\partial \ln L(\mathbf{B} \mid \mathbf{X})}{\partial u} \doteq \sum_{j=1}^J \sum_{k=1}^I \int_\theta \frac{\partial P(\mathbf{x}_j \mid \theta, \mathbf{B}, k)}{\partial u} \frac{f(\theta \mid k) f(k)}{P(\mathbf{x}_j \mid \mathbf{B})} d\theta. \tag{10}$$

Followed by the empirical Bayes method and an approximation of integration by summation, denoted by q quadrature points and $A(\theta_q|k)$ conditional weights approximating $f(\theta_q|k)$, Equation 10 for item parameter u_i can be rewritten as

$$\frac{\partial \ln L}{\partial u_i} \doteq \sum_k \sum_q \frac{A(\theta_q \mid k)}{P_{ik}(\theta_q) Q_{ik}(\theta_q)} \frac{\partial P_{ik}(\theta_q)}{\partial u_i} \sum_{j=1}^J [\mathrm{x}_{ij} - P_{ik}(\theta_q)] \mathrm{f}(k) P_i(\theta_q \mid \mathbf{x}_j, k). \tag{11}$$

Since x_{ij} can be either 1 or 0, the right hand side of Equation 11 can be rewritten as

$$\sum_k \sum_q \frac{1}{P_{ik}(\theta_q) Q_k(\theta_q)} \frac{\partial P_{ik}(\theta_q)}{\partial u_i} \mathrm{f(k)} (R_{iqk} - P_{ik}(\theta_q) N_{iqk}) \tag{12}$$

where

$$R_{iqk} = \sum_{\mathrm{j}} \mathrm{x}_{\mathrm{ij}} \frac{P(\mathbf{x}_j | 1, \theta_q, \mathbf{B}, k) A(\theta_q | k)}{P(\mathbf{x}_j|\mathbf{B})} \tag{13}$$

$$N_{iqk} = \sum_{\mathrm{j}} \frac{P(\mathbf{x}_j | 1, \theta_q, \mathbf{B}, k) A(\theta_q | k)}{P(\mathbf{x}_j|\mathbf{B})} \tag{14}$$

and

$$\frac{\partial P_{ik}(\theta_q)}{\partial a_i} = D(\theta_q - b_i)P_{ik}(\theta_q)Q_{ik}(\theta_q)$$

$$\frac{\partial P_{ik}(\theta_q)}{\partial b_i} = -Da_i P_{ik}(\theta_q)Q_{ik}(\theta_q). \tag{15}$$

The matrix of second derivatives can be expressed as

$$\frac{\partial^2 \ln L}{\partial b_i^2} = \mathrm{D}^2 \sum_k \sum_q \mathrm{f(k)}(\theta_q - b_i)^2 N_{iqk} P_{ik}(\theta_q)Q_{ik}(\theta_q) \tag{16}$$

$$\frac{\partial^2 \ln L}{\partial b_i^2} = -b_i^{\,2} \sum_k \sum_q a_i^2 N_{iqk} P_{ik}(\theta_q)Q_{ik}(\theta_q) \tag{17}$$

$$\frac{\partial^2 \ln L}{\partial a_i \partial b_i} = \mathrm{D}^2 \sum_k \sum_q a_i (\theta_q - b_i)^2 N_{iqk} P_{ik}(\theta_q)Q_{ik}(\theta_q). \tag{18}$$

Once the item parameters are estimated, the estimation of the examinees' proficiency can be carried out using several existing methods, including the maximum likelihood method (MLE), Bayes modal estimates (MAP), and the expected a posterior (EAP) method. The MLE of ability is described by Lord (1980), and MAP and EAP are both described by Bock and Aitkin (1980). Here, the EAP for the typical model with estimator θ_j is

$$\hat{\theta}_j = E(\theta_j \mid \mathbf{x}_j) \doteq \frac{\sum_k \sum_q \theta_q P(\mathbf{x}_j \mid \theta_q) A(\theta_q \mid \mathrm{k})\mathrm{f(k)}}{\sum_k \sum_q P(\mathbf{x}_j \mid \theta_q) A(\theta_q \mid \mathrm{k})\mathrm{f(k)}}. \tag{19}$$

The variance of the EAP estimator is approximately

$$\mathrm{Var}(\hat{\theta}) \doteq \frac{\sum_k \sum_q (\theta_q - \hat{\theta}_j)^2 P(\mathbf{x}_j \mid \theta_q) A(\theta_q \mid \mathrm{k})\mathrm{f(k)}}{\sum_k \sum_q P(\mathbf{x}_j \mid \theta_q) A(\theta_q \mid \mathrm{k})\mathrm{f(k)}}. \tag{20}$$

Thus, the posterior distribution of the proficiency and switching population distributions can be calculated as

$$P(\theta \mid \mathbf{X}, \mathbf{B}, k) \doteq \frac{P(\mathbf{X} \mid \theta, \mathbf{B}, k)}{\sum_q P(\mathbf{X} \mid \theta_q, \mathbf{B}, k)} = \frac{P(\mathbf{X} \mid \theta, \mathbf{B}, k)}{P(\mathbf{X} \mid \mathbf{B}, k)} \tag{21}$$

and

$$P(k \mid \mathrm{X},\mathrm{B}) \doteq \frac{\sum_{q} P(\mathrm{X} \mid \theta_q,\mathrm{B},k)}{\sum_{k}\sum_{q} P(\mathrm{X} \mid \theta_q,\mathrm{B},k)}. \tag{22}$$

The notion of a prior distribution on the item parameters, proficiency distributions, and switching population distribution can be used during the maximization phase. The item parameters, for example, can be viewed as being drawn from a particular distribution; and updating the parameters could be constrained to meet that particular distribution. Similarly, the proficiency distribution can be assumed to be normal at each switch point, including at the last test item. In addition, $E(\theta|k)$ can be constrained to have a specific functional form in relation to the value of k.

Developing psychometric models that incorporate strategy switching is important for a number of reasons: (1) to characterize examinees' strategy use when it is salient; (2) to detect extraneous strategy influences in estimated model parameters; and (3) to provide an opportunity to incorporate partial knowledge of latent classes. The extension of the HYBRID model discussed above attempts to provide a qualitative evaluation of the response strategies used by examinees, and it does this by allowing a closer examination of interaction between the location of the test items and strategy switching patterns. Standard IRT models do not capture this phenomenon and, as a result, can prompt misleading inferences about the proficiencies of the examinees and the properties of the test items.

Fit Indices for the Extended Model

An index of the goodness of fit was needed for this model. For the ideal condition, G^2 could be used. Use of the chi-square test, however, on data derived from a sparse response pattern distribution was not warranted. In light of this, we sought convergent evidence from both the chi-square test and Akaike's (1985; 1987) "information coefficient" (the AIC). The AIC is defined as -2*log-likelihood ratio plus 2*df. Although the chi-square distribution may not be exactly appropriate, the likelihood ratio for nested models was available for examining model fit. Comparing the fit of the two models, such as the 1PL IRT model versus the 2PL IRT model, can be done by examining the improvement in the log-likelihood ratio, while taking into account the number of degrees of freedom expended. However, when competing models are non-nested, the log-likelihood test is less appropriate. In such instances, the AIC can be used.

In this study, then, the LCM class is associated with a unique group of examinees exhibiting strategy switching. At issue is whether the extended HYBRID model, for which classes are suggested via a theory of item performance, can better account for test performance under speeded conditions than can more traditional psychometric models. To examine this issue we applied the extended HYBRID model to simulated data, as well as data derived from two field studies of a standardized, multiple-choice reading comprehension test.

In an effort to demonstrate the efficacy of the HYBRID model for detecting test speededness by identifying strategy switching, we applied the model to three distinct data sets: (1) simulated data representing the performance of 3,000 examinees on a 70 item test; (2) data from a statewide field test (n=5997 of a 40 item reading comprehension test with a fixed time limits; and (3) data from a study of urban university students (n=752) who took a similar 40 item reading comprehension test under varying time limits. The extended HYBRID model was used to analyze both sets of data, and the simulated and actual strategy switch points were mapped onto the structure of the test.

METHOD

Simulation Study

The simulated data set consists of 3000 ability parameters from a standard normal distribution N(0,1), and 70 pairs of item parameters of the standard 2PL IRT model simulated from independent normal distributions, N(1, 0.4) for b, and N(0.0, 0.8) for s. Based on these simulated parameters, a 3000x70 response matrix was generated. Number of simulees switching to random responses were increased in increments of 50 starting with the 51st item. Thus, among 3000 simulees, 2000 did not switch to random responses (k_j=70), while 50 simulees switched to random response at the 51st item (k_j=50), 50 more at the 52nd item (k_j=51), so on till the last (70th) item (k_j=69). Responses of simulees switching to random response were generated based on c_i=0.2.

Three sets of model parameters were estimated for the simulated data: 1) ordinary 2PL IRT parameters (140 parameters were estimated); 2) HYBRID model parameters (140+60+29=229 parameters); and 3) ordinary 2PL IRT parameters with random responses treated as not presented (140 parameters). For the HYBRID model parameter estimation, $f(\theta|k)$ was constrained to be normal and considered for the last 30 items, making 60 parameters to be estimated. In addition, 29 multinomial parameters were to be estimated to represent f(k), each parameter representing proportion of examinees switching at a particular point. The rationale for this was that if item parameters were estimated using the third option, the IRT item parameter estimation would depend only on the portion of the data that correspond to the IRT model, hence the estimation error would be minimized. This would also mean that less data would be available to estimate item parameters for the last 20 items, thus less accurate estimation would be expected. The IRT item parameters estimates based on the competing models would be computed with the estimates of this third option.

Field Study 1

Data for the first field study come from a statewide administration of a 47-item multiple-choice type reading comprehension test administered to nearly 6000 examinees (N=5997) in the Fall of 1990. The reading comprehension test was comprised of 47 multiple choice items based on 10 reading passages of varying lengths. The test was administered in 50 minutes. Seven items--items 26 to 32--were designated as experimental items and were omitted from the analysis. The examinees, all of whom were enrolled in a large public university system, were both ethnically and linguistically diverse and included roughly 64% whites, 6% Asian Americans, 15% African Americans, and 15% Hispanics. Approximately 11% of the sample were identified as students for whom English was a second language (ESL).

The extended HYBRID model was used in the analysis of these data in an attempt to map the switch points on to the structure of the reading comprehension test. Since this particular reading comprehension test contained a number of brief reading passages followed by a short series of multiple-choice items, mapping the points where examinees were affected by speededness of the test "switch" to random response patterns could be achieved. Applying the extended HYBRID model to these field trial data, then, permitted us to test the utility of model for detecting the effects, if any, of speededness for different groups of examinees.

Field Study 2

The second field study was conducted to further examine the utility of the extended HYBRID model for analyzing test data and detecting speededness for different supgroups of the test taking population-- English as a second language (ESL) and English as a primary language (EPL) examinees--when tests are administered under different time conditions. In the second field study, a quasi-experimental research design was used, in which students from a large public urban university (n= 752) took the test in groups of 30 that were randomly assigned to either a 45-minute or a 60-minute time condition. The sample was ethnically diverse, consisting of 40% African

Americans, 17% Hispanics, 12% Asian Americans, 12% Whites, and 19% unclassified. Moreover, 16% were self-identified as students for whom English was a second language.

Using a strategy similar to Study 1, the extended HYBRID model was used to analyze of these data in an attempt to map the switch points on to the structure of the reading comprehension test under the two administrative conditions. Again, since this test contained a number of brief reading passages followed by a short series of multiple choice items, mapping the points where examinees affected by speeded "switch" to random response patterns could be achieved. Applying the extended HYBRID model to these field trial data, then, permitted us to test the utility of the model for detecting the effects, if any, of the different time limits on both test speededness and the ability estimates for different subgroups of examinees.

RESULTS

Simulation Study

We discuss the results of the simulations from the perspective of the accuracy of the item parameter and ability parameter estimates. As noted earlier, three sets of model parameters were estimated: (1) ordinary 2PL IRT parameters ; (2) HYBRID model parameters; and (3) ordinary 2PL IRT parameters with random responses treated as not presented. Table 1 shows the three sets of estimated item parameters and the values of the model fit indices (i.e., -2*log-likelihood and the AIC). The first 39 items in the item sequence are not included since there were virtually identical among results from three methods. Summary statistics comparing the estimated item and ability parameters are presented in Table 2.

INSERT TABLE 1 and 2 HERE

To further examine the amount and direction of the bias in the item parameter estimates of the competing models, we plotted the item parameter estimates produced by both the IRT only model and the HYBRID model against the estimated item parameters based on method 3. These plots are presented in Figure 1.

INSERT FIGURE 1 HERE

The IRT only model overestimated the values of the location parameters for the last 20 items, while the HYBRID model produced much less biased estimates. Similarly, when we compare the plots for the item slope estimates, we see that the IRT only model consistently underestimated the value of the a parameter for the last 20 items. Again, the HYBRID model produced less biased estimates of the location parameters for the last 20 items. When the accuracy of estimated parameters for the last 10 items is considered, the IRT only model is clearly inaccurate.

Accuracy of the estimates of the ability parameter is, perhaps, more important when test data are believed to be influenced by speededness. Careful inspection of the ability estimates by the IRT only method suggests that this method leads to biased estimates because ability for the first 2000 simulated examinees is slightly over estimated, while the θ's for the remaining 1000 simulated examinees are underestimated. Figure 1 presents a plot of ability estimates by IRT only method against the ability estimates by the method three, distinguishes the overestimations and underestimations of θ when standard IRT models are applied to speeded test data. Figure also included a plot of the HYBRID model θ estimates against the method three. There is a clear reduction in the bias estimates of ability under the HYBRID model.

Analysis of the mean deviation and the root mean square deviations (RMSD's) of the ability estimates produced when the methods were used to analyze the data suggests that the HYBRID model produced significant improvements over the IRT only model. Table 2 indicates that the RMSD of the ability estimates is nearly twice as large for the IRT only model for those in sequence between 2501 through 3000. Severity of estimation bias is greater for those with higher ability, more specifically for simulees with θ greater than 1.

The estimated proportion of simulees who switched to random responding strategy,f(k)'s are presented in Table 1, with the clear demarcation of the 51st item where random responding started. Slight overestimation of f(k)'s before 50th item (about 1%) and near the 70th item (about 3%) were found. The cumulative distribution of switched population is presented in Figure 2a. It shows that the estimate closely follow the real distribution indicated by the solid line.

INSERT FIGURE 2a HERE

The competing models were also evaluated in terms of their ability to identify lack of speededness of the test when appropriate. This was done by using the original IRT only data before random responses replaced part of the data. The two methods of the IRT only and the HYBRID model produced nearly identical results in estimated parameters and model fit, and the cumulative distribution of switched proportion only amounted to 0.4% at the last item. The cumulative distribution of switched population is presented in Figure 2b. Comparison of two cumulative distributions of switched simulees presented in Figures 2a and 2b makes it easy to identify which data set contains the speeded simulees. In addition, the HYBRID model correctly identified the data as not speeded in this simulated data set, and estiamated IRT parameters remain unbiased.

INSERT FIGURE 2b HERE

Field Study

When analyzing the field study data for speededness, we relied on the traditional approach of observing the performance of the various subgroups of examinees on items appearing three-fourths of the way through the test. Since items 26-32 of the .est booklet were omitted from our analyses, the item sequence was re-labeled 1 through 40. Item 30, therefore, represents the point at which three-fourths of the test has been completed. Table 3 presents the proportions of the various subgroups, across items, classified as belonging to the "switch" group.

INSERT TABLE 3 HERE

At the point where 75% of the test is completed, 21% of Blacks, 18% of Hispanics, and 17% of Asian were identified as having switched to a random response strategy. These proportions are dramatic when contrasted with the fact that only 8% of the White examinees apparently switched response strategies at this juncture. Similarly, 18% of the ESL and 11% of the EPL subgroups switched to a random response strategy at this point in the test.

As expected, the increase in speededness continues as we extend our analysis to the point at which 87.5% of the test was completed, i.e., item 35 in Table 3. At this point, 29% of the Blacks have switched to a random response strategy, compared to 12% of Whites. Unexpectedly, we found that Asians were affected by speededness of the test. The previous analysis using only omitted responses did not indicate any evidence of speededness among Asians. In fact these data include similar proportions of omitted responses for Whites and Asians. Moreover, nearly 25% of the ESL group switched to a random responding when attempting the items linked to the next to last reading passage. Figure 3 presents the increase in the cumulative proportions of the three minority groups compared to White examinees.

INSERT FIGURE 3 HERE

Field Study 2

In the second field study, we set out to examine the effects of differing time limits on test speededness. The analysis focused on the cumulative proportions of EPL and ESL examinees who switched to a random response strategy in the 45-minute and 60-minute test conditions. Table 4 shows the cumulative proportions of the switching groups over the last twenty test items.

INSERT TABLE 4 HERE

Looking particularly at the distributions after item 36, the last item associated with the penultimate reading passage, we see, as expected, that both testing time and linguistic competence seem to affect the switching strategy. In general, those examinees who had 60 minutes to complete the test had a somewhat lower rate of switching to a random response strategy, 19% for the 60 minute group versus 25% for the shorter, 45 minute time condition. The difference is greater among EPL students, 26% vs. 16%. Contrary to our expectation, the shortened testing time seemed not to affect strategy switching for the ESL examinees. Across groups, however, the organizaton and structure of the reading test is captured well by the cumulative proportions in each of the strategy switching categories presented in Table 2. We see, for example, that there are two points where the cumulative proportions show marked changes, at the 34th and 37th items. These two items correspond well to the structure of the test itself, since both are the first items in the testlets associated with the test's last two reading passages. Figure 4 presents the increase in the cumulative proportions of the ESL and non-ESL populations affected by speededness.

INSERT FIGURE 4 HERE

CONCLUSION

The extended HYBRID model appears to be a promising method for estimating the effects of test length and testing time on test speededness. The analyses of the simulated data, where the item and ability parameters were known, suggest that the model was robust to both a "switching" and an "omitting" speededness strategy. The two field studies revealed much the same thing. The extended HYBRID model tended to characterize examinees' strategy use when it was salient. Perhaps more importantly, it detected the extraneous influences of strategy switching in latter portions of a test on the estimated model parameters, both item parameters and ability parameters.

When we used this model to map the salient strategy switches onto the structure of multiple-choice reading comprehension tests in our field studies, we found that strategy switching followed patterns closely related to the testlet structure of the tests. Examinees tended to switch their response strategies more abruptly between items that are structurally linked to test passages near to the end of the test as expected by classical models of test speededness.

Moreover, the model pointed up the potential differences in test speededness for members of different ethnic and linguistic groups. These differences, however, may be more salient on tests of reading comprehension where surface features of language play a prominent role in the test content. Further research is needed to shed light on the important issue of how test speededness affects the scores of various subgroups on tests that vary in content and purpose. This research focused on tests of reading comprehension, and the model worked relatively well. It remains to be seen whether this model-based approach to detecting test speededness will work as well on tests of mathematical ability or vocabulary where the comprehension demands would be expected to affect performance less.

REFERENCES

Aitkin, M. (1989). Direct likelihood inference. *Unpublished manuscript.* Isreal: Tel Aviv University.

Akaike, H. (1985). Prediction and entropy. In A.C. Atkinson & S.E. Fienberg (Eds.) *A celebration of statistics* (pp. 1-24). NY: Springer-Verlag.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317-332.

Bejar, I.I. (1985). Test speededness under number-right scoring: An analysis of the Test of English as a Foreign Language. *ETS Research Report (RR-85-11).* Princeton, NJ: Educational Testing Service.

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

Cronbach, L.J., & Warrington, W.G. (1951). Time-limit tests: Estimating their reliability and degree of speeding. *Psychometrika, 16(2),* 167-188.

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

Donlon, T.F. (1980). An annotated bibliography of studies of test speededness. *GRE Board Research Report (GREB 76-9R).* Princeton, NJ: Educational Testing Service.

Gulliksen, H. (1987). *Theory of mental tests.* Hillsdale, NJ:. Erlbaum Associates.

Gitomer, H. & Yamamoto, K. (1991). Performance modeling that integrates latent trait and class theory. *ETS Research Report (RR-91-1).* Princeton, NJ: Educational Testing Service.

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

Haertel, E.H. (1984b). Detection of a skill dichotomy using standardized achievement test items. *Journal of Educational Measurement*, 21, 59-72.

Haertel, E.H. (1991, in press). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement.*

Hambleton, R., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Boston, MA: Klumwer-Nijhoff Publishing.

Harwell, M.R., Baker, F.B., & Zwarts, M. (1988). Item parameter estimation via marginal maximum likelihood and an EM algorithm: A didactic. *Journal of Educational Statistics, 13* (3), 234-271.

Hulin, C.L., Drasgow, F. & Parsons, C.K. (1983). *Item response theory: Application to psychological measurement.* IL: Dow Jones-Irwin.

Kendall, M.G., & Stuart, A. (1967). *The Advanced Theory of Statistics* (Vol.2). NY: Hafner.

Kyllonen, P.C., Lohman, D.F., & Snow, R.E. (1984). Effects of aptitudes, strategy training, and task facets on spatial task performance. *Journal of Educational Psychology, 76(1),* 130-145.

Kyllonen, P.C., Lohman, D.F., & Woltz, D.J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology, 76(4)*, 1325-1345.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum Associates.

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

Mislevy, R. J. (1983). Item response models for grouped data. *Journal of Educational Statistics*, 8, 271-288.

Mislevy, R.J. (1991, in press). Foundations of a new test theory. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests*. Hillsdale, NJ: Erlbaum.

Mislevy, R.J. and Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, 55(2), 195-215.

Mislevy, R.J., Yamamoto, K, & Anacker, S. (1991, in press). Toward a test theory for assessing student understanding. In R.A. Lesh (Ed.), *Assessing higher-level understanding in middle-school mathematics*. Hillsdale, NJ: Erlbaum.

Nunnally, J.C. (1978). *Psychometric theory*. NY: McGraw-Hill, Inc.

Rindler, S.E. (1979). Pitfalls in assessing test speededness. *Journal of Educational Measurement, 16*(4), 261-270.

Schmitt, A.P., Dorans, N.J., Crone, C.R, & Maneckshana, B.T. (1991). Differential speededness and item omit patterns on the SAT. *ETS Research Report (RR-91-50)*. Princeton, NJ: Educational Testing Service.

Secolsky, C. (1989). Accounting for random responding at the end of the test in assessing speededness on the Test fo English as a Foreign Language. *ETS Research Report (RR-89-11)*. Princeton, NJ: Educational Testing Service.

Snow, R.E., and Peterson, P. (1985) Cognitive analyses of tests: Implications for redesign. In S. Embretson (Ed.), T*est design: Developments in psychology and psychometrics*. NY: Academic Press.

Stafford, R.E. (1971). The speededness quotient: A new descriptive statistic for tests. *Journal of Educational Measurement, 8*(4), 275-278.

Swineford, F. (1956). Technical manual for users of test analyses. *ETS Statistical Report (SR 56-42)*. Princeton, NJ: Educational Testing Service.

Swineford, F. (1974). The test analysis manual. *ETS Statistical Report (SR-74-06)*. Princeton, NJ: Educational Testing Service.

Wise, L.L. (1986). *Latent trait models for partially speeded tests*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.*Report (RR-89-41)*. Princeton, NJ: Educational Testing Service.

Yamamoto, K. (1987). *A model that combines IRT and latent class models*. [Unpublished Doctoral Dissertation] Champaign-Urbana, IL: University of Illiniois.

Yamamoto, K. (1989). *Hybrid model of IRT and latent class models.* Research Report. Princeton, NJ: Educational Testing Service.

Yamamoto, K. (1990). HYBILm: A computer program to estimate the HYBRID model. Princeton, NJ: Educational Testing Service.

Yamamoto, K. & Gitomer, D. (1993). Application of a HYBRID model to a test of cognitive skill representation. In N. Frederiksen, I. Bejar, & R. Mislevy (Eds.) *Test Theory for a New Generation of Tests.*

List of Tables

List of Figures

TABLE 1. Estimated Item Parameters in the Simulated Data Set

	1 IRT only		2 HYBRID model			3 IRT only with not-presented	
Item	a	b	a	b	f(k)	a	b
41	1.15	-2.78	1.11	-2.89	0.0	1.17	-2.72
42	1.39	-1.59	1.38	-1.65	0.0	1.42	-1.57
43	0.68	0.27	0.68	0.22	0.1	0.69	0.27
44	0.71	1.59	0.74	1.51	0.0	0.74	1.54
45	0.93	0.23	0.94	0.18	0.1	0.95	0.23
46	0.50	1.55	0.51	1.48	0.1	0.51	1.51
47	1.22	-0.52	1.22	-0.56	0.1	1.24	-0.51
48	0.95	-0.43	0.95	-0.49	0.1	0.97	-0.43
49	0.73	1.06	0.75	1.00	0.3	0.76	1.03
50	0.91	-0.15	0.92	-0.20	0.3	0.93	-0.14
51	1.14	0.96	1.21	0.89	0.8	1.23	0.93
52	0.61	-0.89	0.65	-1.00	1.9	0.66	-0.91
53	0.72	2.49	1.01	2.08	1.8	0.99	2.15
54	0.92	1.26	1.08	1.11	1.5	1.08	1.17
55	0.62	-0.45	0.72	-0.63	1.2	0.72	-0.56
56	0.92	0.18	1.08	0.04	1.1	1.10	0.08
57	0.54	0.88	0.61	0.68	1.7	0.61	0.72
58	0.92	1.14	1.22	0.93	1.6	1.25	0.99
59	0.54	3.18	1.15	2.22	1.6	1.15	2.26
60	0.78	1.56	1.10	1.28	2.2	1.09	1.33
61	0.23	-2.62	0.43	-2.73	1.6	0.44	-2.59
62	0.69	1.27	0.86	1.01	1.7	0.88	1.05
63	0.68	1.14	0.96	0.84	1.7	0.93	0.88
64	0.41	0.14	0.57	-0.41	1.9	0.58	-0.33
65	0.50	-0.91	1.14	-1.27	1.8	1.16	-1.18
66	0.34	3.71	0.73	2.26	2.0	0.68	2.43
67	0.72	0.17	1.47	-0.25	2.0	1.46	-0.19
68	0.56	2.12	1.33	1.44	2.3	1.31	1.47
69	0.24	-2.06	1.01	-2.31	2.1	1.09	-2.01
70	0.33	-0.92	0.81	-1.59	1.0	0.90	-1.41
Fit of the model							
-2*log-likelihood[1] No. of Parameters AIC	189,731 140 190,011		185,348 229 185,806			not Comparable	

1 The -2*log-likelihood for the IRT model on the omit data is not comparable due to the fact that 300*10 responses were never included in calculating the likelihood; hence, it was not reported here.

TABLE 2. Fit of the model and accuracy of estimated parameters[2] by the HYBRID.model and the 2PL IRT model

RMSD of item parameter estimates against option (3)				
Parameters	(1) IRT only		(2) HYBRID	
	Mean Deviation	RMSD	Mean Deviation	RMSD
Items 1-70				
Slope b	.02	.22	.02	.03
Location s	.09	.24	-.06	.08
Items 1-50				
Slope b	.02	.03	.02	.02
Location s	.00	.03	-.06	.06
Items 51-60				
Slope b	.21	.25	.00	.01
Location s	.22	.33	-.05	.06
Items 61-70				
Slope b	.48	.51	.01	.05
Location s	.39	.53	-.11	.14
Ability (simulees)				
1-2000	.03	.06	-.00	.05
2001-2500	-.03	..08	-.01	.07
2501-3000	-.10	.21	-.00	.09
2501-3000, θ<.0	.02	.08	-.02	.06
2501-3000, .0<θ<1	-.02	.07	-.02	.05
2501-3000, 1<θ<2	-.34	.25	.04	.10
2501-3000, 2<θ	-1.0	.61	.22	.19

2 Deviation was calculated using the following formula; for the slope parameter, deviation = 1- estimate (by methods 1 or 3)/ estimate (by method 3), and for the location and ability parameters, deviation = estimate(by methods 1 or 2)-estimate(by method 3). The RMSD was calculated using the above values.

TABLE 3. Cumulative Proportion of "Switching " Response Patterns by Examinee Subgroup

Item	Race/Ethnicity				Language		Total
	Asian	Black	Hispanic	White	ESL	EPL	
N	348	920	514	3819	405	5359	5997
20	.00	.00	.00	.00	.00	.00	.00
21	.06	.06	.07	.02	.06	.03	.04
22	.07	.07	.08	.02	.07	.04	.04
23	.07	.07	.08	.02	.07	.04	.04
24	.07	.07	.08	.02	.07	.04	.04
25	.09	.10	.10	.03	.09	.05	.06
26	.10	.13	.11	.04	.11	.07	.07
27	.10	.13	.11	.05	.12	.07	.07
28	.12	.15	.12	.05	.13	.08	.08
29	.14	.17	.15	.06	.15	.09	.10
30	.14	.18	.15	.06	.15	.09	.10
31	.17	.21	.18	.08	.18	.11	.12
32	.20	.25	.22	.10	.21	.14	.14
33	.20	.25	.22	.10	.21	.14	.14
34	.21	.26	.23	.11	.22	.15	.15
35	.24	.29	.26	.12	.25	.17	.18
36	.24	.29	.26	.12	.25	.17	.18
37	.26	.31	.28	.14	.27	.19	.20
38	.28	.34	.31	.16	.29	.21	.22
39	.28	.34	.31	.16	.30	.21	.22
40	.34	.38	.36	.27	.36	.30	.31

TABLE 4. The Cumulative Proportion of EPL and ESL Examinees in the Strategy Switching Groups for Each Time Condition.

Item	45 minutes			60 minutes		
#	All	EPL	ESL	All	EPL	ESL
20	.00	.00	.00	.00	.00	.00
21	.00	.01	.02	.00	.00	.01
22	.00	.01	.03	.00	.01	.01
23	.00	.01	.03	.02	.01	.02
24	.03	.02	.06	.04	.02	.05
25	.06	.05	.09	.06	.04	.09
26	.06	.06	.09	.07	.05	.10
27	.08	.08	.11	.08	.06	.11
28	.08	.08	.11	.08	.06	.11
29	.09	.09	.11	.08	.06	.11
30	.09	.09	.12	.09	.06	.11
31	.09	.09	.12	[illegible]	.06	.11
32	.10	.09	.12	.09	.07	.12
33	.10	.10	.12	.09	.07	.12
34	.16	.16	.21	.14	.11	.19
35	.17	.17	.22	.14	.11	.19
36	.17	.17	.22	.14	.11	.19
37	.25	.26	.25	.19	.16	.23
38	.29	.30	.27	.22	.20	.26
39	.29	.31	.27	.22	.20	.27
40	.30	.32	.27	.23	.21	.27

Figure 1: Comparison of Estimated IRT parameters for the IRT only and the HYBRID models

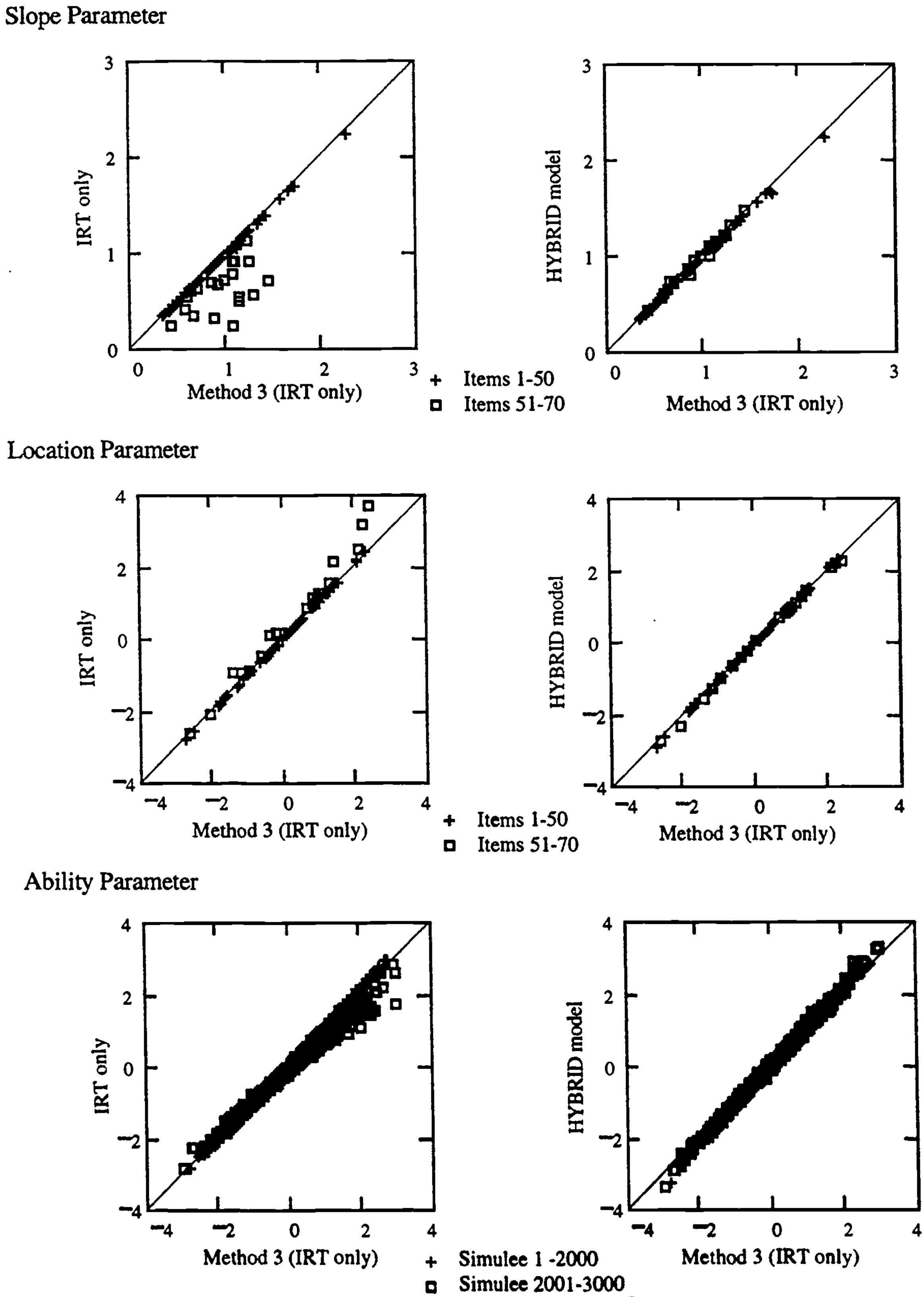
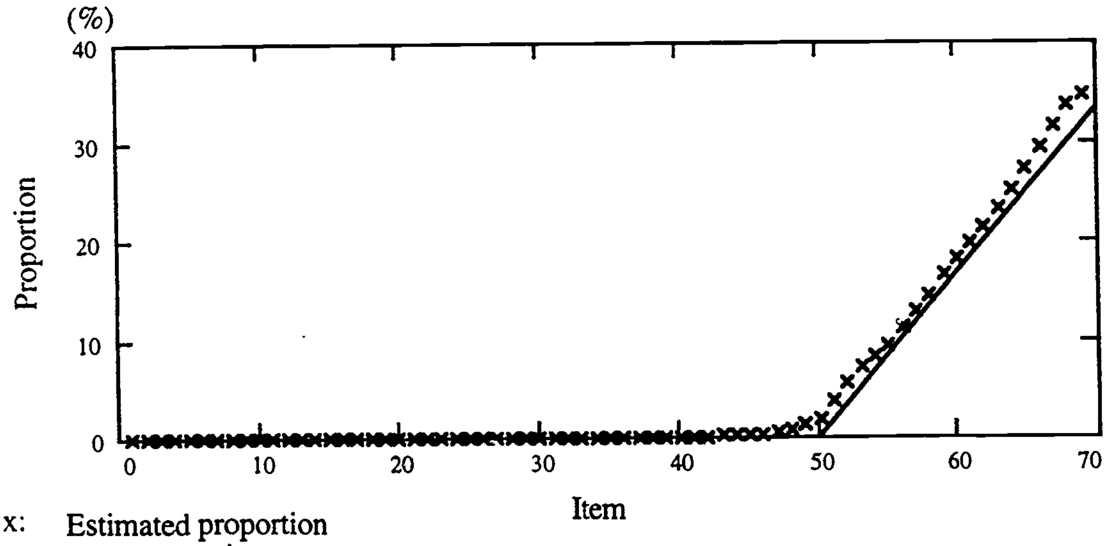
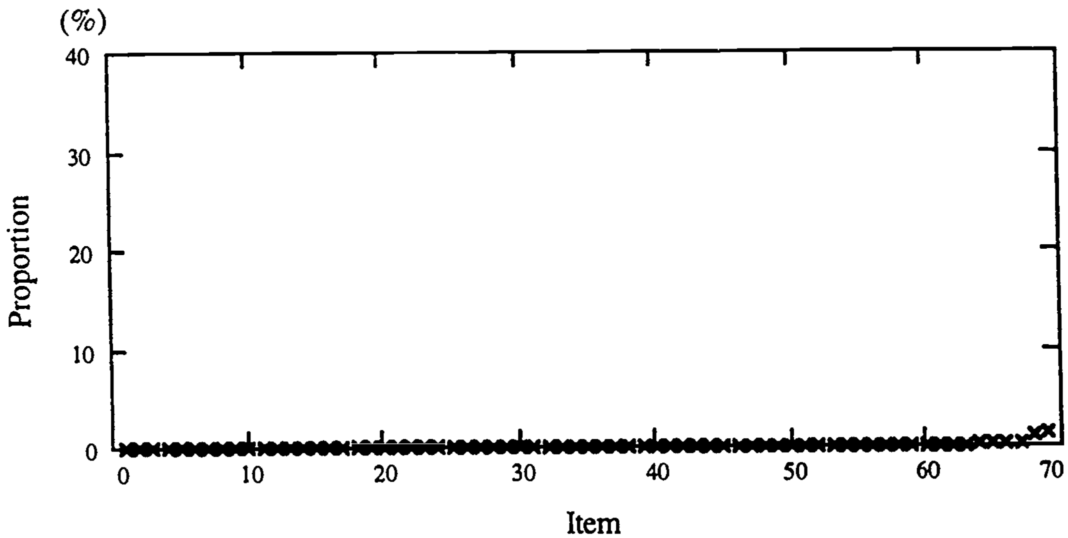


FIGURE 2a. Estimated and true cumulative proportions of switched to random response population of a simulated speeded data.



x: Estimated proportion
_: True proportion

FIGURE 2b. Estimated and true cumulative proportions of switched to random response population of a simulated not speeded data.



x: Estimated proportion

FIGURE 3. Estimated cumulative proportion of switched to random response population by ethnicity of a reading comprehension test.

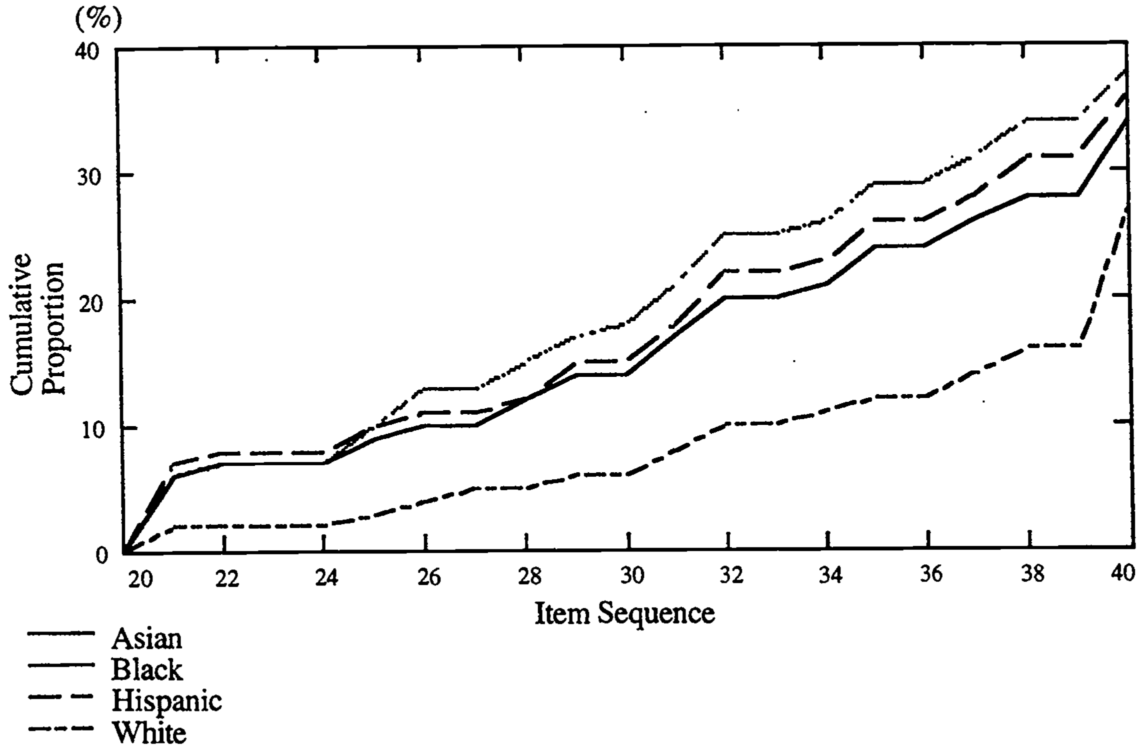


FIGURE 4. Estimated cumulative proportion of switched to random response population for ESL and EPL examinees for the 60 minutes time limit condition.