A New Multi-Class Mixture Rasch Model for Test Speededness

Andrew A. Mroch, Daniel M. Bolt, James A. Wollack

University of Wisconsin-Madison

e-mail: aamroch@wisc.edu

Paper Presented at the Annual Meeting of the National Council on Measurement in Education, April 2005, Montreal, Quebec

Abstract

A multi-class mixture Rasch model (MMRM) is proposed to account for test speededness. The model (1) allows speededness effects to emerge at different test locations for different examinees and (2) accounts for speededness effects due to rushed responses as opposed to strictly random guessing. This model is examined using data sets simulated from a recently proposed speededness model (Wollack & Cohen, 2004) and compared against two previously proposed models: the two-class mixture Rasch model (Bolt, Cohen, & Wollack, 2002) and the HYBRID model (Yamamoto, 1987; Yamamoto & Everson, 1997). Results for the MMRM appeared to be very similar to those observed for the two-class mixture Rasch model. For the datasets generated in this study, the HYBRID model appeared to most closely recover the latent class mixing proportions and item parameters, although it tended to overestimate the effects of speededness on the end-of-test items.

A New Multi-Class Mixture Rasch Model for Test Speededness

Test speededness effects are often observed when examinees do not have sufficient time to finish a test. When examinees are rushed or run out of time, they often fail to adequately answer items at the end of the test. One implication of speededness is its adverse effects on item response theory (IRT) parameter estimates. For example, item difficulty parameters for items at the end of the test may be overestimated compared to their difficulties if administered earlier (Bolt, Cohen, & Wollack, 2002; Oshima, 1994). Several models have been proposed to address problems regarding speededness, including a two-class mixture Rasch model (MRM; Bolt et al., 2002) and the HYBRID model (Yamamoto 1987; Yamamoto & Everson, 1997). These models address speededness effects through the introduction of latent examinee classes that are distinguished by individual differences in speededness effects.

The purpose of this paper is to propose a new multi-class mixture Rasch model (MMRM) for isolating the effects of test speededness, to compare this model against previous speededness models, and to investigate its performance using several illustrative data sets simulated from a recently proposed simulation model for speededness data (Wollack & Cohen, 2004). The proposed MMRM blends the appealing features of the MRM and HYBRID models (described below). Specifically, the MMRM (1) allows speededness effects to emerge at different test locations for different examinees, and (2) accounts for speededness effects due to rushed responses as opposed to only random guesses.

Multi-class Mixture Rasch Model

Several models that address speededness were examined in this study. The speededness model introduced here is a type of multi-class mixture Rasch model (MRMM; Rost, 1990). In the MMRM, multiple latent classes are distinguished by the end-of-test item locations at which their responses become speeded (if at all). The MMRM assumes that examinees belonging to the same latent class experience common item difficulties for items at the end of the test. By assuming the effects of speededness can be captured by changes in the item difficulty parameter of an item, the model allows examinee ability to remain relevant to item responses that are speeded, representing the effect of a rushed rather than random response.

Under the MMRM, the probability that an examinee *j* answers an item *i* correctly is written as follows:

$$P(u_{ij} = 1 | q_{gj}, b_{ig}, g) = \exp(q_{gj} - b_{ig}) / \left[1 + \exp(q_{gj} - b_{ig})\right], \qquad (1)$$

where

u_{ij} is the 0/1 response of examinee *j* to item *i* (0 = incorrect response, 1 = correct response),

q_{gj} is the latent ability parameter of examinee *j* in class *g* {*g* = 1, 2, …, *k* latent classes}, and

b_{ig} is the difficulty parameter for item *i* in class *g*.

Equation 1 is similar to the equation for the Rasch model, the key difference being subscript *g*, which indexes each latent class. The MMRM allows different Rasch model estimates to exist for each latent class; in the current application, only end-of-test item parameters were allowed to differ across classes, and constraints were applied to force the classes to represent classes distinguished by speededness.

Table 1 illustrates the latent class profiles of an example MMRM application involving 40 items and 7 latent classes. For all examinees, a certain number of items at the beginning of the test are assumed not speeded (procedures for identifying the number of such items are described later). The point at which speeded responses begin to occur distinguishes the latent classes. Note that for all classes, it is assumed that the items are answered in order, such that when a given item response is speeded, responses to all subsequent items are also speeded. For example, in Table 1, item responses for examinees in latent class 3 are speeded for the last two items of the test (and not speeded for the first 38). Equality constraints are placed on the difficulty parameters for all unspeeded items across classes (for latent class 3, this corresponds to the first 38 items) as well as for all speeded items across classes. Further, an ordinal constraint is applied to each speeded item defined in each class, such that the speeded item difficulty is always higher (that is, more difficult) than the difficulty for the same item in the nonspeeded case.

Finally, in order to avoid a confounding between the latent ability (q) and classes (g), a norming condition, $\sum_i b_{ig} = 0$, was applied to the difficulty parameters within each class (Rost, 1990). Application of this norming condition is consistent with typical applications of the model and allows classes to be defined more by an item score profile than by number correct. As a result, we assume an examinee's response pattern exhibits effects of speededness when the relative difficulties of items at the end of the test are higher than the relative difficulties of items at the beginning.

Consequently, there exist two difficulty parameters for each potentially speeded item, one for its speeded condition and one for its nonspeeded condition. One difficulty parameter exists for each item that is nonspeeded across all classes (e.g., items at the beginning of the test). In addition, mixing proportions, p_g, indicate the proportion of examinees in each latent class, where $\sum_g p_g = 1$. Finally, each class is associated with a m_{q_g} parameter, representing the mean of q_g, examinee ability, in that class (the variance of q_g for all examinee classes was set to 1). In the current application, however, the m_{q_g} were constrained as a function of the relative difficulties of speeded and nonspeeded items in each class; that is, $m_{q_{speed_g}} = m_{b_{nonspeed}} - m_{b_{speed_g}}$. This was done to keep ability unassociated with class membership. Alternatively, it could be assumed that these are parameters to be estimated; however, practical experience suggests that due to the relatively low numbers of examinees in each of the speeded classes, the parameters are not estimated very well, and their presence can deleteriously affect how the classes are defined.

Two competing models to the MMRM are considered next.

Two-class Mixture Rasch Model

An alternative to the MMRM is the two-class mixture Rasch model (MRM) proposed by Bolt et al. (2002). The MRM is a special case of the MMRM described above, where the number of latent classes is equal to two. The equation for this model follows from Equation 1, where g = 2. Instead of a single unspeeded latent class and multiple speeded latent classes, however, the two-class MRM defines a single speeded latent class and a single nonspeeded latent class and thus does not explicitly account for the different locations in the test at which speededness may begin. Similar to the MMRM, a constraint is placed on the difficulties within each latent class such that $\sum_i b_{ig} = 0$ and mixing proportions, are estimated. But unlike the MMRM, estimation of the MRM proceeds without the constraint that m_{q_g} be a function of the relative item difficulties, as the mean can generally be well-estimated due to the larger number of examinees in the one speeded class.

Table 2 illustrates the latent class profiles of a two-class MRM with 40 items and 6 potentially speeded items. The 34 items at the beginning of the test are not speeded for either class, while the 6 items at the end of the test are treated as speeded for the speeded class. This differs from Table 1, which displays the latent class profiles for a similar test using the MMRM, where multiple speeded latent classes are defined. Similar to the MMRM, equality constraints are placed on the difficulties for the items at the beginning of the test, where speededness is assumed not to be present, and an ordinal constraint is applied to each speeded item defined in the speeded class, such that the speeded item difficulty is always higher than the nonspeeded item difficulty. Despite the presence of only one speeded class, it was anticipated that the two-class model could still accommodate conditions in which speeded examinees vary in terms of the item location where speededness begins. More is said on this issue later.

HYBRID Model

The third model considered in this study is Yamamoto's HYBRID model (Yamamoto 1987; Yamamoto & Everson, 1997). The HYBRID model assumes that for each examinee that is speeded, there is a point on the test at which the examinee switches response strategy from attempting to solve the items to guessing randomly among the item alternatives. Similar to the MMRM, the HYBRID model associates different latent classes of examinees with different item locations at which test speededness first begins, thereby allowing examinees to become speeded at different points on a test. Examinee performance on the test is modeled by an item response model up to the point where random guessing begins, and as a random Bernoulli trial thereafter. (In this study, a Rasch model was assumed for the item response model, but it is possible to use other models, for an example, see Bolt, Mroch, & Kim, 2003). Therefore, the latent class profiles listed in Table 1 for the MMRM could also be applied using the HYBRID model. As for the MMRM, speededness is always assumed to occur in an ordered fashion, such that once an examinee becomes speeded, the examinee is speeded on the remainder of the test. The HYBRID model, thus models the probability of an examinee j answering an item i correctly as follows:

$$P(u_{ij}=1\,|\,\theta_j,b_i,s_{ig})=\left(\exp(\theta_j-b_i)/\left[1+\exp(\theta_j-b_i)\right]\right)^{1-s_{ig}}\times\left(\gamma_{ig}\right)^{s_{ig}} \qquad (2)$$

where

u_{ij} is the 0/1 response of examinee j to item i (0 = incorrect response, 1 = correct response),

q_j is the latent ability parameter of examinee j,

b_i is the difficulty parameter for item i,

γ_{gi} is the probability of examinees in class g {g = 1, 2, …, k latent classes} randomly guessing the correct answer to item i, and

s_{ig} identifies whether or not an item is speeded in class g (0 = not speeded, 1 = speeded).

In the current application of the HYBRID model, before items become speeded (before examinees switch to random guessing, where $s_{ig} = 0$), the probability of a correct response is modeled using a Rasch model. When items become speeded (and random guessing is used, where $s_{ig} = 1$), the probability of a correct response is then modeled as a random Bernoulli trial (where the expected probability of a correct response is equal to the inverse of the number of item responses; for example 1/5 = 0.2 for items with 5 alternatives).

Comparisons Among the Three Speededness Models

The three speededness models have several similarities and differences that warrant consideration. First, the specific versions of the speededness models used here all assume the Rasch model as the item response model upon which the unspeeded class of examinees will be based. These can be modified to accommodate other IRT models (e.g., see Bolt et al., 2003 or Rost, 1996), but are assumed here for simplicity. Second, each model uses latent classes to account for examinee differences in the occurrence of speededness effects. Third, each model defines speededness as emerging at the end of the test, and thus assumes a sequential ordering in how examinees respond to the items.

The primary difference between the MMRM and HYBRID models considered is in the nature of the speededness effects assumed. In the MMRM, speededness effects are viewed as making items more difficult; in the HYBRID model, they emerge as random guessing. The MRM defines two latent classes and can be viewed as an approximation to the MMRM.

Estimating Model Parameters Using Markov Chain Monte Carlo

A Markov Chain Monte Carlo (MCMC) algorithm was used to estimate the parameters of each of the three speededness models. WinBUGS 1.4 software (Spiegelhalter, Thomas, & Best, 2003) was used for these purposes. Sample WinBUGS code for estimating the MMRM is provided in Appendix B. Convergence of the MCMC solution was determined by inspecting plots of sampling histories for estimated parameters, and the means of the sampled values (after burn-in) were used as estimates of the parameters. The prior distributions for the MMRM and MRM were as follows:

b_{ig} ~ Normal(0, 1)

q_{gj} ~ Normal(m_g, 1)

m_{q1} ~ Normal (0, 1) (for the unspeeded class)

and $m_{q2} \ldots m_{qk}$ are functions of the unspeeded and speeded item parameters (as defined above for the MMRM only; both m_{q1} and m_{q2} are estimated in the two-class MRM),

c_j ~ Categorical ($\pi_1, \pi_2, \ldots \pi_k$), where $c_j = \{1, 2, \ldots, k\}$ is a class membership parameter, and

($\pi_1, \pi_2, \ldots \pi_k$) ~ Dirichlet (1, …, 1).

For the HYBRID model,

b_i ~ Normal(0, 1)

q_j ~ Normal(0, 1)

$c_j \sim \text{Categorical}\ (\pi_1, \pi_2, \ldots \pi_k)$, and

$(\pi_1, \pi_2, \ldots \pi_k) \sim \text{Dirichlet}\ (1, \ldots, 1)$

For each simulated data set examined in this study, item parameters and latent class mixing proportions were estimated for the three speededness models. In the estimation process, latent class memberships are sampled for each examinee at each stage in the chain, but vary over the course of the chain. Consequently, each examinee can be viewed as having a posterior probability of membership in each class, and is not explicitly assigned to any one class.

A twofold approach was used to report results. First, the estimates of each speededness model were compared to the generating parameters to evaluate the recovery of item parameters and latent class mixing proportions. Second, the three speededness models were compared regarding their effectiveness at recovering end-of-test item parameters, both at the item and subtest score level.

Speededness Simulation

To examine performance of the MMRM, MRM, and HYBRID model and to compare their performances, several illustrative data sets were simulated using a speededness simulator proposed by Wollack and Cohen (2004). This simulator generates data that builds in realistic sources of speededness but that is too complex to estimate. Consequently, all of the speededness models fit in the current study were approximations to the true generating model. The speededness simulator assumes that speededness (a) emerges at the end of a test, (b) emerges at different points at the end of the test for different examinees, and (c) is manifest by an “erosion” in performance as time runs out and whose effects may vary for different examinees. Some examinees will devote the necessary time to most items and guess on only a few, while others will divide their time so they can read and attempt all items but may be somewhat rushed. The model used to generate speeded and unspeeded item responses for examinees is as follows for item i and examinee j:

$$P(u_{ij}=1|q_j,c_i,a_i,b_i,h_j,l_j)=c_i+(1-c_i)\left\{\frac{e^{a_i(q_j-b_i)}}{1+e^{a_i(q_j-b_i)}}\cdot\min(1,[1-(\tfrac{i}{n}-h_j)])^{l_j}\right\}, \quad (3)$$

where

n is the number of items on the test,

c_i, a_i, b_i, and q_j correspond to the 3-parameter IRT model pseudo-guessing, discrimination, difficulty, and ability parameters, respectively,

h_j $(0 \le h \le 1)$ is the speededness point parameter,

l_j $(l_j \ge 0)$ is the speededness rate parameter, and

min(x, y) is the smaller of the two values x and y.

Notice that everything to the left of the min() term is a 3-parameter IRT model, which by itself reflects an unspeeded item. However, the min() term in this equation builds a speeded component into the model; if this term is less than 1, it reflects erosion in the probability that an examinee correctly responds to the item. When the min() term has a value of 1 (when the item is not speeded), Equation 3 reflects an unspeeded 3-parameter IRT model. At the other extreme, if the min() term is 0, the equation reduces to a guessing model where the probability of a correct response is 1/(number of response alternatives).

The point on the test where this erosion occurs depends on the η_j parameter, which identifies for each examinee the proportion of unspeeded items on the test. For example, an examinee with η_j of .90 on a 40-item test means that speededness effects emerge 90% of the way through the test, or at item 36. In Equation 3, examinee responses at item n x η_j (where n is the number of items on the test) and thereafter are modeled by a 3-parameter IRT model multiplied by the component that erodes (or decreases) the probability that a person gets the item correct. The amount of this erosion depends on two things: (1) the parameter λ_j and (2) the number of items away from the speededness starting point (n x η_j) that a given item is located. Once an examinee's responses become speeded, the term $\frac{i}{n} - h_j$ in Equation 3 is raised to power λ_j, which controls how quickly the probability of a correct response decreases. Also, the farther past the speededness point an examinee gets on the test, the larger $\frac{i}{n} - h_j$ becomes, which leads to a larger reduction in the probability of a correct response.

Four data sets were simulated to illustrate estimation of the MMRM, evaluate its parameter recovery, and to compare its results against the two-class MRM and HYBRID models. Each data set consisted of 40 items, and assumed the following distributions for the speededness point and rate parameters: h_j ~ beta (20, 2), l_j ~ lognormal (3.912, 1). The distribution of the η_j parameter as beta (20, 2) was selected because (a) most values tended to be between .7 and 1.0, reflecting speededness that starts between 70% and close to 100% of the way through the test, and (b) the mean of this distribution, .91, corresponds to 9%, or roughly four items at the end of a 40 item test being speeded. Observed distributions of where speededness started across speeded examinees for simulated data sets indicated that six items at the end of the test appeared speeded, which was then used as the number of items considered in defining the latent classes in the speededness models. The distribution of l_j was chosen as lognormal (3.912, 1) to create a moderately strong rate of decline in the probability of correct response for speeded examinees.

Two data sets were simulated with 1000 examinees (300 speeded), and two were simulated with 1500 examinees (500 speeded). Generating item difficulty parameters and examinee ability parameters were drawn from a standard normal distribution, that is: b_i ~ Normal (0, 1) and q_j ~ Normal (0, 1). In the sample size = 1500 condition, the generating parameters for the first 1000 examinees were the same as those for the sample size = 1000 data sets, but an additional 500 examinees were added. Initially, after completing runs with 1000 examinees, model estimates for speeded groups appeared somewhat unstable, so 500 examinees (200 speeded examinees) were added to observe model estimates for slightly larger speeded groups. Therefore, the only difference between the two 1000 examinee data sets and two 1500 examinee data sets was the number of examinees (all other parameters were identical). Each set of data is referred to as “set A” (with 1000 or 1500 examinees) or “set B” (with 1000 or 1500 examinees). Generating parameters for the pseudo-guessing (c_i) and discrimination (a_i) parameters were fixed at 0.2 and 1, respectively, such that the speededness simulator data sets corresponded to a Rasch model with a lower asymptote fixed at .2 and a speededness component. Fixing the generating pseudo-guessing parameter at .2 matched the speeded condition for the HYBRID model, where speededness was modeled as guessing (with probability of .2). Furthermore, fixing the pseudo-guessing parameter to .2 and the discrimination parameters to 1 contributed to a closer balance between (a) realistically simulated data (by including a guessing parameter greater than zero) and (b) model misfit (in the form of deviations from the Rasch model assumed by each of the speededness models used in this study) in addition to that already caused by test speededness.

Model parameters for each data set were estimated via MCMC using each of the three speededness models (MMRM, MRM, and HYBRID). In addition, item parameters were estimated via MCMC for a Rasch model to examine how well this model would recover item parameters for data sets containing speededness but without modeling speededness effects. If end-of-test item parameters were accurately recovered via the Rasch model for a data set containing speededness, a model accounting for speededness may not be warranted. For each model (MMRM, MRM, HYBRID, and Rasch) 15,000 iterations were sampled from the Markov chain and 5,000 iterations were used as burn-in, leaving 10,000 iterations sampled from the posterior distribution to use as estimates of model parameters.

To examine nonspeeded item parameter recovery, comparison of the estimated nonspeeded class item parameters to the nonspeeded generating item parameters from the speededness simulator was of primary interest. To do this, the estimated item parameters were first equated to the generating item parameters using test characteristic curve (TCC) equating (Stocking & Lord, 1983) via the computer program EQUATE (Baker, 1994) to ensure that the item parameter estimates were on the same scale. The first 30 items in each data set were used to link estimated item parameters to the generating item parameters for the full 40-item test. These items were used because we wanted to equate on as pure a subset of items as possible; the first 30 items generally did not contain speededness effects in the simulated data sets.

Model performance was examined in several ways. First, the generating and estimated item parameters were plotted to provide an overall display of results. Second, the generating and estimated latent class mixing proportions were examined to identify how well each speededness model was able to recover the proportion of examinees in each latent class. Third, the TCC for the last six test items, i.e., the expected number of items correct, was plotted to illustrate (a) the expected end-of-test performance under nonspeeded conditions according to each of the models and (b) the extent to which each model was able to "purify" item parameter estimates for the last six items on the test, so that they reflected what the estimates should look like in nonspeeded conditions. By looking at the TCC, it was expected that we could better assess any systematic bias that occurred for each of the models in estimating the parameters for the last six items. Fourth, to summarize the difference in item parameter recovery for each estimated model compared to the generating parameters, an expected standardized difference index (ESDI) was computed for the last six items for each data set. The ESDI quantifies for each item the average squared difference between the estimated probabilities of correct response and true probabilities weighted by the distribution of ability. The equation for ESDI is as follows:

$$ESDI = \frac{1}{\sum w(q)} \times \left(\sum_q \left[\hat{P}_{TCC}(q) - P_{TCC}(q) \right]^2 \times w(q) \right) \quad (4)$$

where,

$w(q)$ is the weight based on the distribution of ability q,

$\hat{P}_{TCC}(q)$ is the estimated number correct score for a given test characteristic curve, and

P_{TCC} is the true generating expected number correct score for a given test characteristic curve.

Results

Figures 1a to 1d display plots of generating item parameters and estimated item parameters for the MMRM, MRM, HYBRID, and Rasch models for each simulated data set. These data are also presented in table form in Appendix A. Recall that the generating item difficulty parameters were randomly drawn from a standard normal distribution. As such, most item difficulties fell within a range of -2 to 2. One peculiarity occurred for data set A, where the last item on the test had a difficulty parameter of almost -4, which means that the item was extremely easy when nonspeeded; more will be said on this shortly.

The item parameter estimates for the first 35 items in each data set were similar (where, within each speededness model, item parameters were fixed to be equal across all examinees) while the last 6 items in each data set differed somewhat across speededness models. In terms of item parameter recovery, the HYBRID model generally appeared to either perform similarly to the MRM and MMRM, or better, especially for the last item. Also, the MMRM and MRM appeared to perform very similarly to each other. Finally, as expected, the Rasch model generally did not appear to be as accurate at recovering item parameters as the MMRM, MRM, and HYBRID models.

Tables 3a to 3d display the generating and model estimated latent class mixing proportions for each of the simulated data sets. For the 1000-examinee data sets, 70% of examinees were simulated to be unspeeded and 30% were simulated to be speeded. For the 1500-examinee data sets, 67% of examinees were simulated to be unspeeded and 33% were simulated to be speeded. Across data sets, the results appeared to be similar for each model. The MMRM identified between 14% and 18% of examinees as speeded, the MRM identified between 15% and 19% of examinees as speeded, and the HYBRID model identified between 24% and 29% of examinees as speeded. Therefore, each of the three speededness models underestimated the proportion of speeded examinees in the simulated data sets, although the HYBRID model most closely recovered the latent class mixing proportions.

Figures 2a to 2d display the generating and estimated unspeeded class test characteristic curves for the last six items of each simulated data set. Across the four simulated data sets, several observations warrant mentioning. First, the Rasch model displayed the largest underestimate of nonspeeded end-of-test scores across the ability scale. This was expected because the Rasch model does not take into account speededness and was thus expected to be most affected by erosion in performance at the end of the test.

Second, once again the MMRM and MRM tended to perform similarly; in this case, both appeared to underestimate the expected end-of-test scores when nonspeeded. For the 1000-examinee data sets, these underestimates started near an ability of 0 and gradually increased in magnitude as ability decreased; thus, the MMRM and MRM showed larger underestimates of end-of-test performance for low ability examinees. A similar pattern of results was observed for the larger sample size.

By contrast, the HYBRID model tended to overestimate end-of-test performance under nonspeeded conditions. For the 1000-examinee data sets, the effects were evident over much of the ability scale, generally beginning at theta of –2 or –1.5 with the largest difference in the range of theta –1 to 1, where the largest number of examinees was located.

Table 4 displays the ESDI for each speededness model and each simulated data set. The ESDI was computed by taking values of ability from –4 to +4 in increments of 0.05. Weights $w(q)$ were computed at each theta level from a standard normal distribution because the generating data were drawn from a standard normal distribution. For the set A data sets, the MRM had the smallest ESDI values (.04 and .04), followed by the MMRM (.08 and .05). The HYBRID model had larger ESDI values (.11 and .17) than the MRM and MMRM. Finally, the Rasch model had the largest ESDI values (.24 and .26). For set B data sets, the HYBRID model now had the smallest ESDI value for the 1000-examinee data set (.03) but for the 1500-examinee data set, the MRM had a smaller ESDI (.04). The MRM and MMRM ESDI values decreased as the size of the speeded group increased (MRM went from .09 to .04 and MMRM went from .11 to .07). The Rasch model again had the largest ESDI values (.37 and .28). Overall, the MRM and MMRM ESDI values appeared to decrease as sample size increased, while the HYBRID model ESDI values increased as sample size increased. Although only based on a small number of analyses, this pattern is not entirely unexpected, as explained in the discussion.

Discussion

In this paper, we presented a Multi-class Mixture Rasch Model as a tool that can be used to address the issue of speededness on paper-pencil tests. Using several data sets simulated to contain speededness, we illustrated the estimation of this model and compared the results to other existing models for speededness. All of the speededness models studied appear to be effective in reducing the influence of speededness on end-of-test item parameter estimates; all recovered generating parameters and probabilities of correct response better than the Rasch model. However, the limited number and scope of the simulated data sets used to estimate the speededness models did not provide much evidence that one speededness model is particularly better than another. More extensive simulation work may be necessary to judge the meaningfulness of the differences observed here.

Despite the limitations of the current study, several interesting findings emerged. First, the MRM and MMRM tended to perform similarly. Initially, it was anticipated that the MMRM might perform slightly better than the MRM because data were simulated to exhibit speededness at different points near the end of the test, but this was not generally the case. The MRM, assuming one speeded class comprised of six speeded items, performed similarly to the MMRM, assuming 6 speeded classes comprised of between 1 and 6 items. This may be due in part to the nature of the Bayesian estimation procedure used, where latent class membership is estimated along with the item parameter estimates and latent class mixing proportions. As noted, each examinee is actually associated with a posterior probability of belonging to each class; hence it may be possible for the two-class model to provide a close enough approximation to make it just as useful as one that explicitly accounts for the location at which speededness occurs. Because we are mainly interested in modeling the item parameter estimates under nonspeeded conditions, a precise account of what occurs for the speeded examinees may not be critical. Bigger differences among the models might be observed if the purpose were to identify individuals for whom a test is speeded, perhaps for purposes of excluding such examinees from the equating process (Wollack, Cohen, & Wells, 2003)

Second, the HYBRID model performed quite well and in many respects better than the MRM and MMRM, which was surprising given that the simulated data included general erosion in performance at the end of the test, and not necessarily to chance levels (though performance did frequently reduce to chance for speeded items). Interestingly, however, although the HYBRID model generally underestimated the mixing proportions for each latent class of examinees, it overestimated the effects of speededness, as evidenced by the TCC differences. Consequently, the better performance of the HYBRID model can probably best be attributed to a cancellation of effects associated with an underestimation of the number of speeded examinees and an overestimation of speededness effects. Of course, the critical question that arises in evaluating this finding is whether the speededness simulator used is actually the manner in which speededness effects occur on real tests. The generating model was chosen because it does have perhaps the most rigorous support as a realistic model for speededness, and is also a model that differed from any of the models used here to remove speededness effects.

Even if this model is realistic, it is clear that more simulation work is needed to evaluate the performance of the estimation methods considered in this paper. For example, alternative values or distributions could be chosen from which to draw values for the speededness parameters of the speededness simulator. In this study, we chose a beta (20, 2) distribution for the speededness point parameter and lognormal (3.912, 1) for the speededness rate parameter. More consideration should be given as to whether these or alternative distributional choices would be more realistic. The logical next step in this project may be to pursue a range of simulation conditions anticipated to affect the three speededness models including such variables as: (1) the magnitude and variability of speededness effects, (2) the difficulty of end-of-test items, and (3) the sample size. The difficulty of end-of-test items would be an interesting condition to examine because of its potential effects on the HYBRID model's assumption that speeded examinees switch to random guessing at the end of the test. Often tests are designed so as to have more difficult items at the end, unlike the very easy item simulated to occur at the end of the set A dataset.

A third practical finding from this study that deserves greater attention is the tendency for all of the models to underestimate the number of speeded examinees. To some degree, this may again be a function of the way in which the generator introduces speededness through a gradual erosion of performance. However, it may also be a consequence of the estimation method, and the use of priors on the mixing proportions. The fact that the proportion of examinees in speeded classes tended to be low may be due to a certain level of bias when a Bayesian method was used to estimate the models. It would be interesting to compare methods when the mixing proportions are fixed at their true values as a way of examining which model comes closer to addressing the nature of speededness introduced by this speededness generator.

Finally, it should be noted that practical applications of these models require specifying the number of items at the end of the test that are speeded. In practice this number of items is not known in advance as was assumed here. Thus, additional examination of methods for adequately specifying the number of end-of-test items that are speeded would be useful. For example, while not included in this study, examination of a nonlinear two factor exploratory factor analysis for several of the data sets used in this study revealed that the last six items at the end of the test loaded on a second factor that could be interpreted as a speededness factor. It may be reasonable to use such analyses as a preliminary way of checking how many items appear to have been affected by speededness.

References

Baker, F. B. (1993). Equate 2.0: A computer program for the characteristic curve method of IRT equating. *Applied Psychological Measurement, 17,* 20.

Bolt, D. M., Cohen, A. S., & Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: Application of a mixture Rasch model with ordinal constraints. *Journal of Educational Measurement, 39,* 331-348.

Bolt, D. M., Mroch, A. A., & Kim, J.-S. (April, 2003). *An empirical investigation of the Hybrid IRT model for improving item parameter estimation in speeded tests.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Oshima, T. C. (1994). The effect of speededness on parameter estimation in item response theory. *Journal of Educational Measurement, 31*, 200 – 219.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271 – 282.

Rost, J. (1996). Logistic mixture models. In W. J. van der Linden and R. K. Hambleton (Eds.): *Handbook of Modern Item Response Theory* (pp. 449-463). New York: Spring.

Spiegelhalter, D., Thomas, A., & Best, N. (2003). *WinBUGS version 1.4 [computer program]*. Robinson Way, Cambridge CB2 2SR, UK: MRC Biostatistics Unit, Institute of Public Health.

Stocking, & Lord, (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201-210.

Wollack, J. A. & Cohen, A. S. (April, 2004). *A model for simulating speeded test data.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Wollack, J. A., Cohen, A. S., & Wells, C. S. (2003). A method for maintaining scale stability in the presence of test speededness. *Journal of Educational Measurement*, *40*, 307-330.

Yamamoto, K. (1987). *A model that combines IRT and latent class models*. Unpublished doctoral dissertation. University of Illinois, Champaign – Urbana.

Yamamoto, K. & Everson, H. (1997). Modeling the effects of test length and test time on parameter estimation using the HYBRID model. In J. Rost and R. Langeheine (Eds.): *Applications of Latent Trait and Latent Class Models in the Social Sciences (pp. 89 – 98)*. New York: Waxman

Table 1
Illustration of latent class profiles for an MMRM speededness model with 40 items and 7 latent classes.*

	Items												
Latent Class	1	2	3	…	32	33	34	35	36	37	38	39	40
1	U	U	U	…	U	U	U	U	U	U	U	U	U
2	U	U	U	…	U	U	U	U	U	U	U	U	S
3	U	U	U	…	U	U	U	U	U	U	U	S	S
4	U	U	U	…	U	U	U	U	U	U	S	S	S
5	U	U	U	…	U	U	U	U	U	S	S	S	S
6	U	U	U	…	U	U	U	U	S	S	S	S	S
7	U	U	U	…	U	U	U	S	S	S	S	S	S

U = Unspeeded item.
S = Speeded item.
* These same latent class profiles apply to the HYBRID model, however the nature of speededness (defined by the "S" items) is modeled differently across the two models.

Table 2
Illustration of latent class profiles for a MRM speededness model with 40 items and a speeded class defined by the last six items on the test.

	Items												
Latent Class	1	2	3	…	32	33	34	35	36	37	38	39	40
1	U	U	U	…	U	U	U	U	U	U	U	U	U
2	U	U	U	…	U	U	U	S	S	S	S	S	S

U = Unspeeded item.
S = Speeded item.

Table 3a
Mixing Proportions for Simulated Data Set A, 1000 Examinees

Speeded Class	MMRM	MRM	HYBRID	Generating*	Final Generating**
1 (unspeeded)	.86	.83	.76	.70	.70
2 (1 speed)	.02		.02		.03
3 (2 speeded)	.04		.05		.07
4 (3 speeded)	.03		.06	.30	.05
5 (4 speeded)	.01		.02		.03
6 (5 speeded)	.02		.06		.05
7 (6 speeded)	.02	.17	.03		.03

*The proportion of examinees simulated to contain speededness was specified at 300 examinees (300/1000 = 30%).
**The proportion of examinees corresponding to each speeded class based on simulation parameters (which was generated from a Beta(20, 2) distribution); note that there were examinees that differed from modeled classes, most having more than 6 items speeded (these accounted for 4% of examinees or 13% of the speeded group).

Table 3b
Mixing Proportions for Simulated Data Set A, 1500 Examinees

Speeded Class	MMRM	MRM	HYBRID	Generating*	Final Generating**
1 (unspeeded)	.82	.81	.73	.67	.67
2 (1 speed)	.03		.03		.03
3 (2 speeded)	.04		.06		.09
4 (3 speeded)	.03		.04	.33	.06
5 (4 speeded)	.03		.05		.04
6 (5 speeded)	.03		.05		.04
7 (6 speeded)	.02	.19	.04		.04

*The proportion of examinees simulated to contain speededness was specified at 500 examinees (500/1500 = 33%).
**The proportion of examinees corresponding to each speeded class based on simulation parameters (which was generated from a Beta(20, 2) distribution); note that there were examinees that differed from modeled classes, most having more than 6 items speeded (these accounted for 3% of examinees or 9% of the speeded group).

Table 3c
Mixing Proportions for Simulated Data Set B, 1000 Examinees

Speeded Class	MMRM	MRM	HYBRID	Generating*	Final Generating**
1 (unspeeded)	.83	.85	.72	.70	.70
2 (1 speed)	.03		.06	.30	.02
3 (2 speeded)	.04		.04		.07
4 (3 speeded)	.03		.05		.04
5 (4 speeded)	.01		.02		.04
6 (5 speeded)	.03		.06		.05
7 (6 speeded)	.02	.15	.04		.02

*The proportion of examinees simulated to contain speededness was specified at 300 examinees (300/1000 = 30%).
**The proportion of examinees corresponding to each speeded class based on simulation parameters (which was generated from a Beta(20, 2) distribution); note that there were examinees that differed from modeled classes, most having more than 6 items speeded (these accounted for 6% of examinees or 20% of the speeded group).

Table 3d
Mixing Proportions for Simulated Data Set B, 1500 Examinees

Speeded Class	MMRM	MRM	HYBRID	Generating*	Final Generating**
1 (unspeed)	.82	.81	.71	.67	.67
2 (speed 1)	.02		.04	.33	.02
3 (speed 2)	.06		.07		.09
4 (speed 3)	.03		.05		.05
5 (speed 4)	.02		.04		.04
6 (speed 5)	.03		.05		.05
7 (speed 6)	.02	.19	.04		.02

*The proportion of examinees simulated to contain speededness was specified at 500 examinees (500/1500 = 33%).
**The proportion of examinees corresponding to each speeded class based on simulation parameters (which was generated from a Beta(20, 2) distribution); note that there were examinees that differed from modeled classes, most having more than 6 items speeded (these accounted for 6% of examinees or 18% of the speeded group).

Table 4
Standardized Difference Index Based on TCCs for Last 6 Items of Simulated Data Sets

Data Set	Number of Examinees*	Model			
		MMRM	MRM	HYBRID	Rasch
Set A	1000	0.08	0.04	0.11	0.24
	1500	0.05	0.04	0.17	0.26
Set B	1000	0.11	0.09	0.03	0.37
	1500	0.07	0.04	0.07	0.28

*1000 examinee data sets had 300 speeded examinees; 1500 examinee data sets had 500 speeded examinees.

Figure 1a
Generating Item Parameters and Item Parameter Estimates, Simulated Data Set A, 1000 Examinees

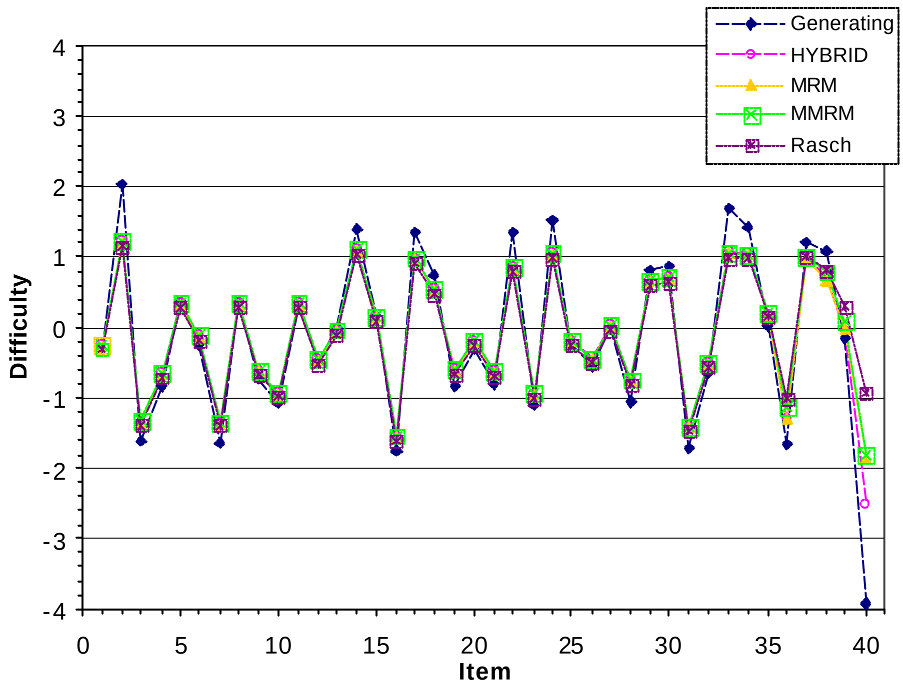


Figure 1b
Generating Item Parameters and Item Parameter Estimates, Simulated Data Set A, 1500 Examinees

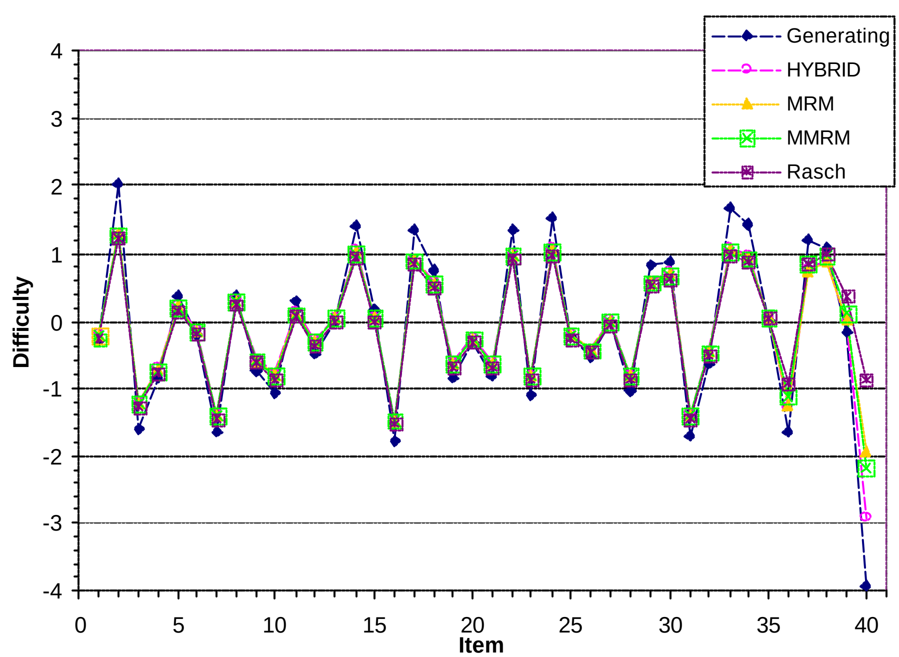


Figure 1c
Generating Item Parameters and Item Parameter Estimates, Simulated Data Set B, 1000 Examinees

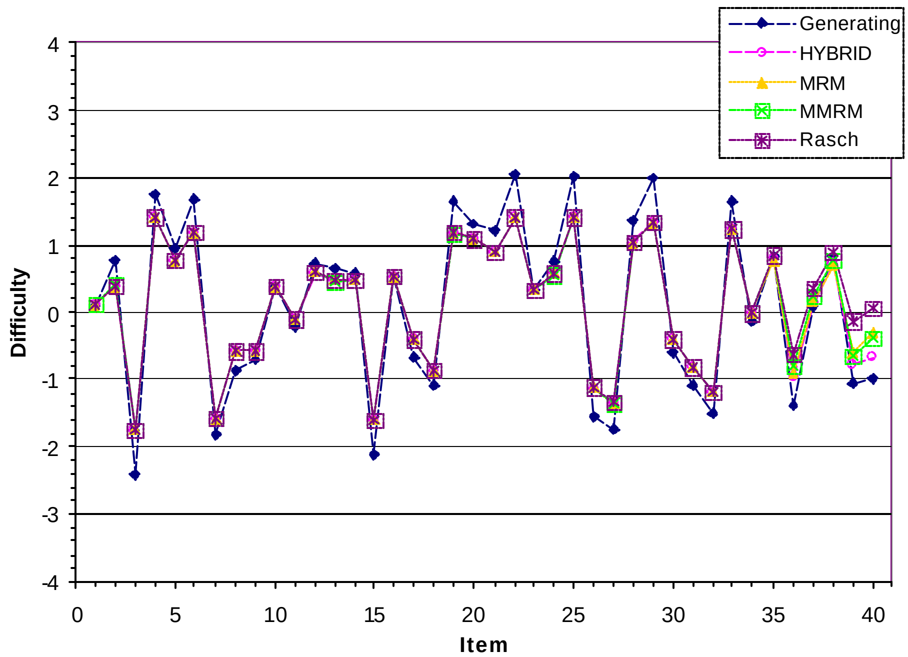


Figure 1d
Generating Item Parameters and Item Parameter Estimates, Simulated Data Set B, 1500 Examinees

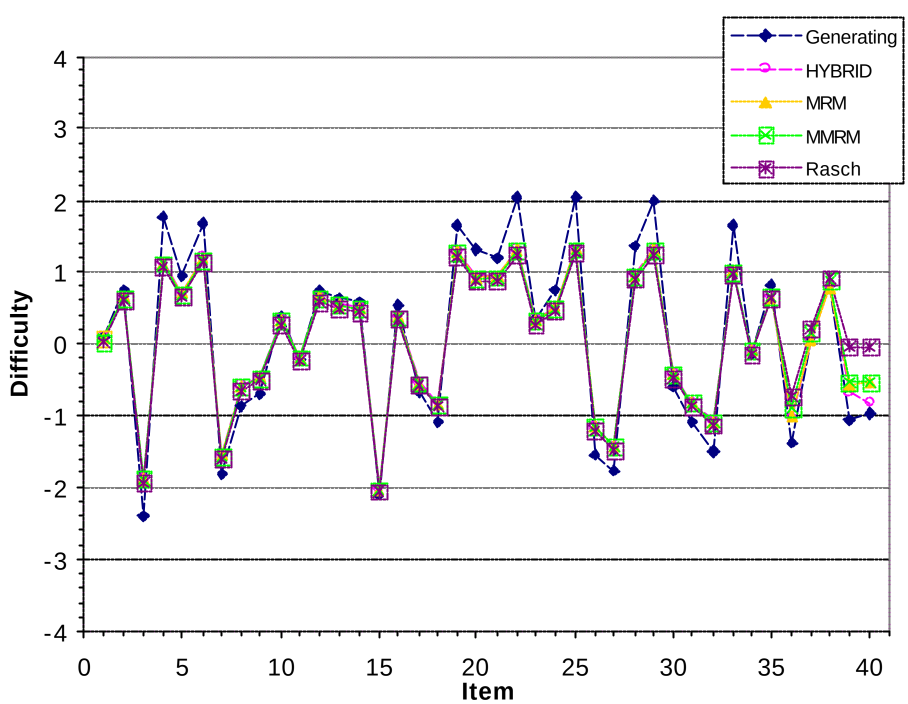


Figure 2a
Test Characteristic Curves for Last Six Items, Simulated Data Set A, 1000 Examinees

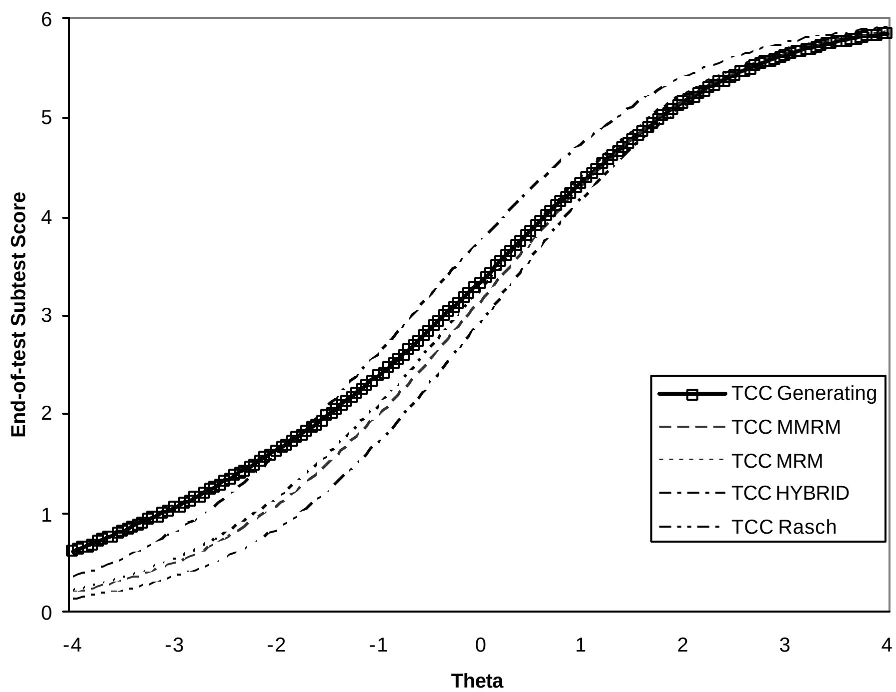


Figure 2b
Test Characteristic Curves for Last Six Items, Simulated Data Set A, 1500 Examinees

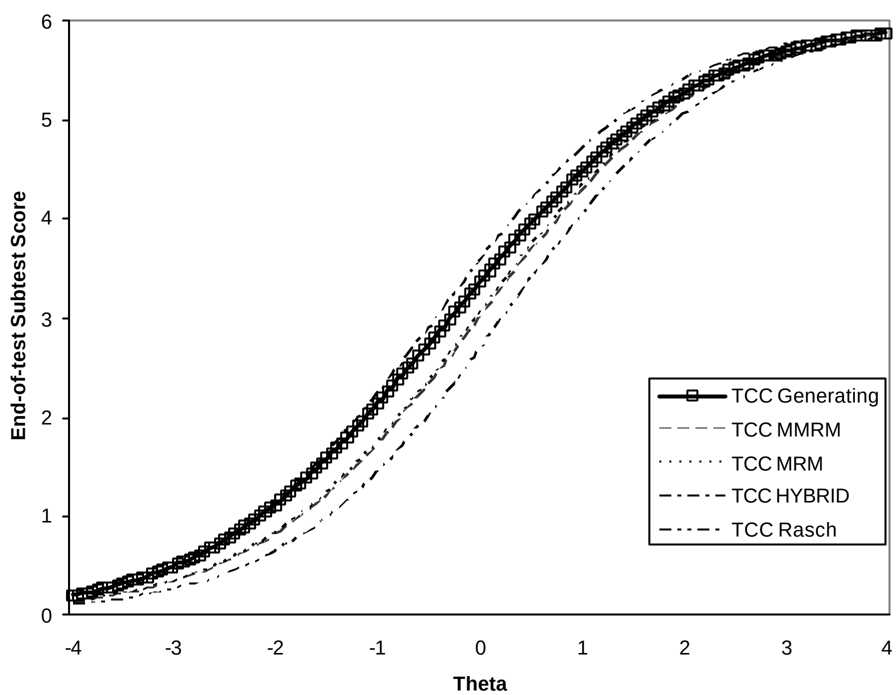


Figure 2c
Test Characteristic Curves for Last Six Items, Simulated Data Set B, 1000 Examinees

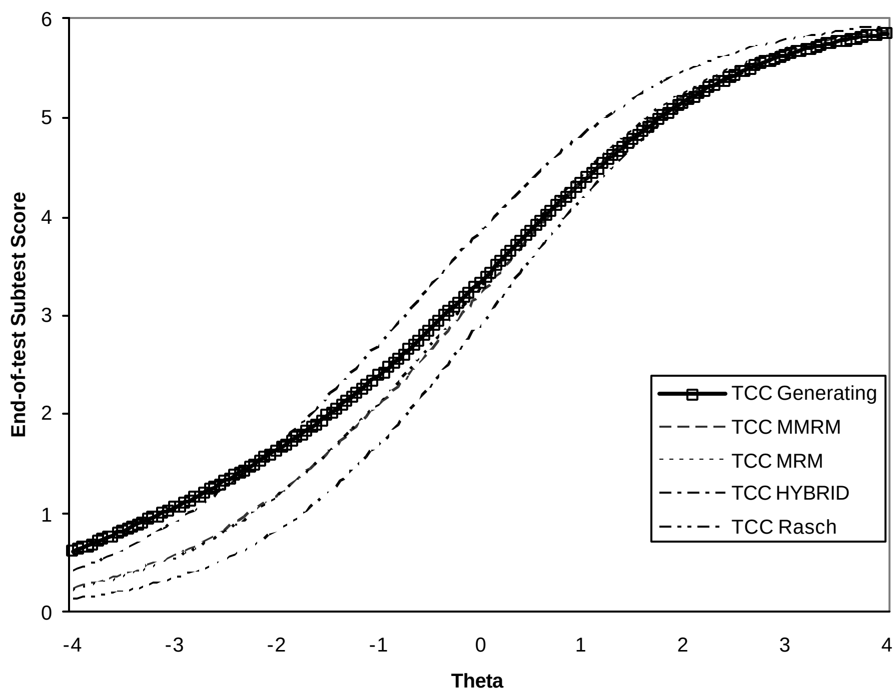
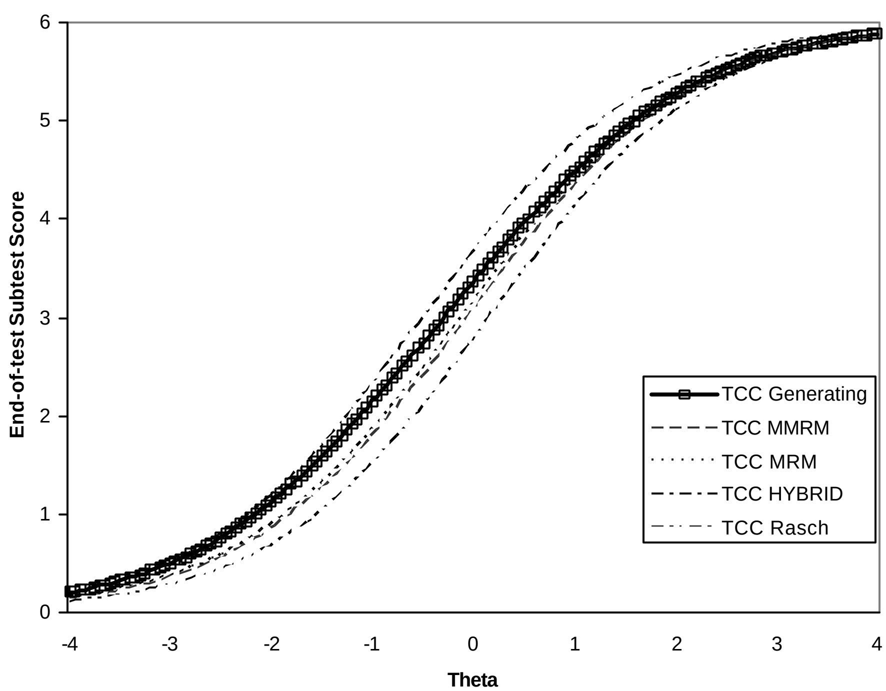


Figure 2d
Test Characteristic Curves for Last Six Items, Simulated Data Set B, 1500 Examinees



Appendix A

Generating Item Parameters and Model Item Parameter Estimates

Data Set A, 1000 Examinees

	Generating	MMRM		MRM		HYBRID		Rasch	
	b	$\hat{b}$	SE	$\hat{b}$	SE	$\hat{b}$	SE	$\hat{b}$	SE
1	-0.30	-0.25	0.07	-0.27	0.07	-0.25	0.08	-0.32	0.08
2	2.02	1.23	0.07	1.21	0.07	1.23	0.08	1.15	0.08
3	-1.61	-1.33	0.09	-1.36	0.09	-1.34	0.09	-1.41	0.09
4	-0.85	-0.65	0.07	-0.67	0.07	-0.65	0.08	-0.72	0.08
5	0.36	0.35	0.07	0.33	0.07	0.35	0.08	0.28	0.07
6	-0.22	-0.12	0.07	-0.13	0.07	-0.12	0.07	-0.19	0.08
7	-1.65	-1.34	0.09	-1.36	0.09	-1.34	0.09	-1.41	0.09
8	0.36	0.35	0.07	0.34	0.07	0.35	0.07	0.28	0.07
9	-0.74	-0.61	0.07	-0.63	0.07	-0.61	0.08	-0.68	0.08
10	-1.08	-0.93	0.08	-0.95	0.08	-0.93	0.08	-1.00	0.09
11	0.28	0.35	0.07	0.33	0.07	0.35	0.07	0.27	0.08
12	-0.48	-0.46	0.07	-0.48	0.07	-0.46	0.08	-0.53	0.08
13	-0.01	-0.04	0.07	-0.06	0.07	-0.04	0.07	-0.12	0.08
14	1.40	1.11	0.07	1.09	0.07	1.11	0.08	1.03	0.08
15	0.17	0.16	0.07	0.15	0.07	0.17	0.07	0.09	0.08
16	-1.77	-1.54	0.09	-1.57	0.09	-1.55	0.10	-1.62	0.10
17	1.36	0.98	0.07	0.96	0.07	0.98	0.07	0.90	0.08
18	0.75	0.55	0.07	0.53	0.07	0.55	0.07	0.48	0.07
19	-0.85	-0.60	0.07	-0.62	0.07	-0.60	0.08	-0.67	0.08
20	-0.31	-0.19	0.07	-0.20	0.07	-0.19	0.08	-0.26	0.08
21	-0.82	-0.62	0.07	-0.64	0.07	-0.62	0.08	-0.70	0.08
22	1.36	0.86	0.07	0.85	0.07	0.87	0.07	0.79	0.08
23	-1.10	-0.94	0.08	-0.97	0.08	-0.95	0.09	-1.02	0.09
24	1.52	1.05	0.07	1.03	0.07	1.06	0.08	0.97	0.08
25	-0.23	-0.17	0.07	-0.19	0.07	-0.17	0.08	-0.25	0.08
26	-0.55	-0.44	0.07	-0.45	0.07	-0.43	0.08	-0.51	0.08
27	-0.01	0.04	0.07	0.02	0.07	0.04	0.07	-0.03	0.08
28	-1.05	-0.75	0.08	-0.77	0.08	-0.75	0.08	-0.82	0.08
29	0.82	0.67	0.07	0.66	0.07	0.67	0.07	0.60	0.07
30	0.86	0.71	0.07	0.69	0.07	0.71	0.07	0.63	0.07
31	-1.71	-1.41	0.09	-1.43	0.09	-1.41	0.09	-1.48	0.10
32	-0.65	-0.49	0.07	-0.51	0.07	-0.49	0.08	-0.57	0.08
33	1.69	1.07	0.07	1.05	0.07	1.08	0.08	0.99	0.08
34	1.42	1.04	0.07	1.03	0.07	1.05	0.08	0.97	0.08
35	0.01	0.21	0.07	0.14	0.08	0.18	0.08	0.15	0.07
36	-1.66	-1.13	0.09	-1.31	0.10	-1.31	0.11	-1.02	0.09
37	1.21	1.02	0.07	0.96	0.08	1.00	0.08	1.01	0.08
38	1.08	0.79	0.07	0.67	0.08	0.71	0.08	0.81	0.08
39	-0.17	0.08	0.08	-0.02	0.09	-0.03	0.09	0.31	0.07
40	-3.93	-1.81	0.25	-1.86	0.19	-2.51	0.32	-0.94	0.08

Data Set A, 1500 Examinees

	Generating	MMRM		MRM		HYBRID		Rasch	
	b	$\hat{b}$	SE	$\hat{b}$	SE	$\hat{b}$	SE	$\hat{b}$	SE
1	-0.30	-0.22	0.06	-0.22	0.06	-0.20	0.06	-0.27	0.06
2	2.02	1.28	0.06	1.29	0.06	1.31	0.06	1.22	0.06
3	-1.61	-1.23	0.07	-1.23	0.07	-1.21	0.07	-1.28	0.07
4	-0.85	-0.73	0.06	-0.72	0.06	-0.70	0.07	-0.78	0.07
5	0.36	0.19	0.06	0.21	0.05	0.23	0.06	0.14	0.06
6	-0.22	-0.15	0.06	-0.15	0.06	-0.13	0.06	-0.20	0.06
7	-1.65	-1.39	0.07	-1.40	0.07	-1.38	0.08	-1.45	0.08
8	0.36	0.29	0.06	0.30	0.06	0.32	0.06	0.24	0.06
9	-0.74	-0.58	0.06	-0.58	0.06	-0.56	0.07	-0.63	0.07
10	-1.08	-0.81	0.06	-0.80	0.06	-0.78	0.07	-0.86	0.07
11	0.28	0.10	0.06	0.11	0.06	0.13	0.06	0.05	0.06
12	-0.48	-0.32	0.06	-0.31	0.06	-0.30	0.06	-0.37	0.06
13	-0.01	0.03	0.06	0.04	0.06	0.05	0.06	-0.02	0.06
14	1.40	1.01	0.06	1.02	0.06	1.04	0.06	0.95	0.06
15	0.17	0.03	0.06	0.04	0.06	0.05	0.06	-0.02	0.06
16	-1.77	-1.48	0.08	-1.49	0.08	-1.48	0.08	-1.54	0.08
17	1.36	0.90	0.06	0.91	0.06	0.93	0.06	0.85	0.06
18	0.75	0.54	0.05	0.55	0.06	0.56	0.06	0.48	0.06
19	-0.85	-0.64	0.06	-0.64	0.06	-0.62	0.07	-0.69	0.07
20	-0.31	-0.27	0.06	-0.26	0.06	-0.24	0.06	-0.32	0.06
21	-0.82	-0.65	0.06	-0.65	0.06	-0.63	0.07	-0.70	0.07
22	1.36	0.97	0.06	0.98	0.06	1.00	0.06	0.92	0.06
23	-1.10	-0.83	0.06	-0.83	0.06	-0.81	0.07	-0.88	0.07
24	1.52	1.03	0.06	1.04	0.06	1.06	0.06	0.97	0.06
25	-0.23	-0.22	0.06	-0.21	0.06	-0.19	0.06	-0.27	0.06
26	-0.55	-0.43	0.06	-0.42	0.06	-0.41	0.06	-0.48	0.06
27	-0.01	-0.01	0.06	0.00	0.06	0.01	0.06	-0.06	0.06
28	-1.05	-0.83	0.06	-0.83	0.06	-0.81	0.07	-0.88	0.07
29	0.82	0.57	0.05	0.58	0.06	0.60	0.06	0.52	0.06
30	0.86	0.68	0.06	0.69	0.06	0.71	0.06	0.62	0.06
31	-1.71	-1.40	0.07	-1.40	0.07	-1.38	0.08	-1.45	0.08
32	-0.65	-0.47	0.06	-0.47	0.06	-0.45	0.06	-0.52	0.06
33	1.69	1.03	0.06	1.04	0.06	1.06	0.06	0.98	0.06
34	1.42	0.93	0.06	0.95	0.06	0.96	0.06	0.88	0.06
35	0.01	0.05	0.06	0.03	0.06	0.03	0.07	0.03	0.06
36	-1.66	-1.13	0.07	-1.26	0.08	-1.28	0.09	-0.93	0.07
37	1.21	0.84	0.06	0.74	0.07	0.77	0.07	0.84	0.06
38	1.08	0.97	0.06	0.87	0.07	0.90	0.07	1.00	0.06
39	-0.17	0.12	0.06	0.02	0.07	0.00	0.08	0.37	0.06
40	-3.93	-2.19	0.24	-1.93	0.16	-2.91	0.36	-0.86	0.07

Data Set B, 1000 Examinees

	Generating	MMRM		MRM		HYBRID		Rasch	
	b	$\hat{b}$	*SE*	$\hat{b}$	*SE*	$\hat{b}$	*SE*	$\hat{b}$	*SE*
1	0.09	0.10	0.07	0.11	0.07	0.11	0.07	0.10	0.07
2	0.76	0.39	0.07	0.39	0.07	0.39	0.08	0.38	0.07
3	-2.41	-1.75	0.10	-1.75	0.10	-1.75	0.11	-1.75	0.11
4	1.76	1.40	0.07	1.41	0.07	1.40	0.08	1.39	0.08
5	0.94	0.77	0.07	0.78	0.07	0.78	0.07	0.77	0.07
6	1.68	1.18	0.07	1.19	0.07	1.19	0.08	1.18	0.08
7	-1.82	-1.57	0.09	-1.57	0.09	-1.57	0.10	-1.58	0.10
8	-0.87	-0.57	0.07	-0.56	0.07	-0.56	0.08	-0.57	0.08
9	-0.70	-0.57	0.07	-0.56	0.07	-0.57	0.08	-0.57	0.08
10	0.39	0.37	0.07	0.38	0.07	0.38	0.07	0.37	0.07
11	-0.22	-0.11	0.07	-0.11	0.07	-0.11	0.08	-0.11	0.08
12	0.73	0.59	0.07	0.60	0.07	0.59	0.08	0.59	0.07
13	0.64	0.46	0.07	0.47	0.07	0.47	0.08	0.46	0.07
14	0.57	0.47	0.07	0.48	0.07	0.48	0.08	0.46	0.07
15	-2.12	-1.60	0.10	-1.60	0.10	-1.60	0.10	-1.60	0.10
16	0.53	0.53	0.07	0.54	0.07	0.54	0.08	0.53	0.07
17	-0.67	-0.40	0.07	-0.40	0.07	-0.40	0.08	-0.40	0.08
18	-1.10	-0.87	0.08	-0.86	0.08	-0.86	0.09	-0.87	0.08
19	1.65	1.17	0.07	1.18	0.07	1.18	0.08	1.17	0.08
20	1.32	1.09	0.07	1.10	0.07	1.10	0.08	1.08	0.08
21	1.2	0.88	0.07	0.89	0.07	0.89	0.07	0.88	0.08
22	2.04	1.41	0.07	1.42	0.07	1.42	0.08	1.40	0.08
23	0.33	0.32	0.07	0.33	0.07	0.33	0.08	0.32	0.07
24	0.74	0.56	0.07	0.57	0.07	0.57	0.07	0.56	0.07
25	2.03	1.40	0.07	1.41	0.07	1.41	0.08	1.40	0.08
26	-1.56	-1.12	0.08	-1.12	0.08	-1.13	0.09	-1.12	0.09
27	-1.76	-1.35	0.09	-1.35	0.09	-1.35	0.10	-1.35	0.10
28	1.36	1.05	0.07	1.05	0.07	1.06	0.08	1.04	0.08
29	2.00	1.34	0.07	1.34	0.07	1.34	0.08	1.33	0.08
30	-0.61	-0.41	0.07	-0.41	0.07	-0.41	0.08	-0.42	0.08
31	-1.08	-0.81	0.08	-0.81	0.08	-0.81	0.09	-0.82	0.08
32	-1.51	-1.18	0.08	-1.18	0.08	-1.18	0.09	-1.18	0.09
33	1.65	1.23	0.07	1.24	0.07	1.24	0.08	1.23	0.08
34	-0.14	-0.02	0.07	-0.01	0.07	-0.02	0.08	-0.02	0.08
35	0.82	0.83	0.07	0.78	0.08	0.81	0.08	0.83	0.08
36	-1.39	-0.79	0.09	-0.92	0.10	-0.97	0.11	-0.62	0.08
37	0.08	0.23	0.07	0.16	0.08	0.15	0.09	0.34	0.07
38	0.82	0.77	0.08	0.70	0.08	0.72	0.09	0.90	0.07
39	-1.06	-0.64	0.12	-0.60	0.11	-0.77	0.13	-0.15	0.08
40	-0.98	-0.38	0.15	-0.30	0.10	-0.66	0.20	0.05	0.08

Data Set B, 1500 Examinees

	Generating	MMRM		MRM		HYBRID		Rasch	
	b	$\hat{b}$	SE	$\hat{b}$	SE	$\hat{b}$	SE	$\hat{b}$	SE
1	0.09	0.08	0.05	0.08	0.06	0.10	0.06	0.04	0.06
2	0.76	0.63	0.05	0.65	0.05	0.66	0.06	0.60	0.06
3	-2.41	-1.89	0.09	-1.89	0.09	-1.88	0.09	-1.93	0.09
4	1.76	1.12	0.06	1.12	0.06	1.14	0.06	1.08	0.06
5	0.94	0.68	0.05	0.69	0.05	0.71	0.06	0.65	0.06
6	1.68	1.17	0.06	1.19	0.06	1.21	0.06	1.14	0.06
7	-1.82	-1.57	0.08	-1.56	0.08	-1.54	0.08	-1.60	0.08
8	-0.87	-0.61	0.06	-0.60	0.06	-0.58	0.06	-0.64	0.07
9	-0.7	-0.48	0.06	-0.46	0.06	-0.45	0.06	-0.50	0.06
10	0.39	0.31	0.06	0.32	0.05	0.34	0.06	0.28	0.06
11	-0.22	-0.21	0.06	-0.20	0.06	-0.18	0.06	-0.24	0.06
12	0.73	0.63	0.05	0.64	0.05	0.66	0.06	0.59	0.06
13	0.64	0.52	0.05	0.53	0.05	0.55	0.06	0.49	0.06
14	0.57	0.47	0.05	0.48	0.06	0.50	0.06	0.44	0.06
15	-2.12	-2.03	0.09	-2.03	0.09	-2.02	0.10	-2.06	0.09
16	0.53	0.37	0.05	0.38	0.05	0.40	0.06	0.34	0.06
17	-0.67	-0.57	0.06	-0.55	0.06	-0.54	0.07	-0.59	0.07
18	-1.1	-0.85	0.06	-0.84	0.06	-0.83	0.07	-0.88	0.07
19	1.65	1.25	0.06	1.27	0.06	1.28	0.06	1.22	0.06
20	1.32	0.90	0.06	0.92	0.06	0.94	0.06	0.87	0.06
21	1.2	0.91	0.06	0.93	0.06	0.95	0.06	0.88	0.06
22	2.04	1.28	0.06	1.30	0.06	1.31	0.06	1.25	0.06
23	0.33	0.31	0.05	0.32	0.05	0.34	0.06	0.28	0.06
24	0.74	0.49	0.05	0.50	0.05	0.52	0.06	0.46	0.06
25	2.03	1.30	0.06	1.32	0.06	1.34	0.06	1.27	0.06
26	-1.56	-1.17	0.07	-1.16	0.07	-1.15	0.07	-1.20	0.07
27	-1.76	-1.44	0.07	-1.43	0.07	-1.42	0.08	-1.47	0.08
28	1.36	0.93	0.06	0.95	0.06	0.97	0.06	0.90	0.06
29	2	1.29	0.06	1.30	0.06	1.32	0.06	1.25	0.06
30	-0.61	-0.45	0.06	-0.44	0.06	-0.42	0.06	-0.48	0.06
31	-1.08	-0.84	0.06	-0.83	0.06	-0.82	0.07	-0.87	0.07
32	-1.51	-1.10	0.07	-1.09	0.07	-1.07	0.07	-1.13	0.07
33	1.65	1.00	0.06	1.01	0.06	1.03	0.06	0.97	0.06
34	-0.14	-0.11	0.06	-0.10	0.06	-0.09	0.06	-0.14	0.06
35	0.82	0.64	0.06	0.62	0.06	0.65	0.06	0.63	0.06
36	-1.39	-0.90	0.07	-1.02	0.09	-1.03	0.09	-0.73	0.07
37	0.08	0.14	0.06	0.05	0.07	0.05	0.07	0.23	0.06
38	0.82	0.86	0.06	0.78	0.07	0.81	0.07	0.93	0.06
39	-1.06	-0.54	0.10	-0.59	0.10	-0.68	0.11	-0.04	0.06
40	-0.98	-0.55	0.11	-0.54	0.10	-0.83	0.17	-0.04	0.06

Appendix B

Sample WinBUGS Code for the MMRM

```
model
{
for (i in 1:NE){
for (j in 1:NI){
p[i,j]<- exp((theta[i]-b[gmem[i],j]))/(1+exp((theta[i]-b[gmem[i],j])))
}}

for (i in 1:NE){
for (j in 1:NI){
r[i,j]~dbern(p[i,j])
}}

for (j in 1:34){
beta[j]~dnorm(0.,1.)
betas[j]<-beta[j]
}

for (j in 35:40){
beta[j]~dnorm(0.,1.)
betas[j]~dlnorm(0,0.25)
}

for (i in 1:NE){
theta[i]~dnorm(mu[gmem[i]],1.)
gmem[i]~dcat(pi[1:7])
}

pi[1:7]~ddirch(alphat[])
mu[1]~dnorm(0,1)

for (i in 2:7){
mu[i]<-mean(betat[1,1:NI])-mean(betat[i,1:NI])
}
for (i in 1:7){
for (j in 1:NI){
betat[i,j]<-beta[j]+(speed[i,j])*betas[j]
}}

for (i in 1:7){
for (j in 1:NI){
b[i,j]<-betat[i,j]-mean(betat[i,1:NI])
}}
}

list(NE=1000, NI=40,alphat=c(1,1,1,1,1,1,1),speed=
structure(.Data=c(
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1), .Dim=c(7,40)),
r=structure(.Data=c(
0,0,1,0,0,1,1,1,1,1,0,1,0,1,1,1,0,0,0,1,1,1,0,1,1,1,0,0,1,1,1,0,0,0,0,0,0,1,0,
1,0,1,1,1,1,0,1,0,1,0,0,1,0,0,1,0,1,0,1,1,0,1,0,1,0,0,1,0,1,1,1,0,1,0,1,0,0,1,0,
.
.
.
1,0,0,1,0,0,1,0,0,1,1,1,0,1,0,1,0,1,1,0,0,0,1,0,0,1,0,0,0,0,1,1,0,0,1,1,0,1,1,1,
0,1,1,1,1,1,1,0,1,1,1,1,1,1,1,1,0,1,1,1,1,1,1,0,1,1,1,1,0,1,1,1,0,1,1,1,1,1,1,1), .Dim=c(1000,40)))
```