A Mixture Rasch Model with Item Response Time Components

J. Patrick Meyer

James Madison University

Paper to be presented at the annual meeting of the National Council on Measurement in Education, New York. March, 2008.

Abstract

Examinees who rapidly respond to test items produce item response patterns and item response time distributions that are distinct from those produced by examinees who do not rapidly guess the answer. These qualitative differences in examinee test-taking behavior bias parameter estimates for item response models that do not control for such behavior. A mixture Rasch model (MRM) with item response time components was proposed and evaluated through a simulation study and application to real test data. Results of the simulation study indicated that the parameters were recovered well. The analysis of extant data indicated that a two-class solution fit better than a one-class solution, and that 15% of examinees engaged in rapid-guessing behavior

A Mixture Rasch Model with Item Response Time Components

Examinees who rapidly respond to test items produce item response patterns and item response time distributions that are distinct from those produced by examinees who do not rapidly guess the answer. The reasons for engaging in rapid-guessing behavior include an examinee's attempt to answer all items on a time-limit test as well as an examinee's endeavor to avoid items that require too much effort. In either case, the response pattern and item response time distributions of rapid-guessing examinees are qualitatively different from examinees who engage an item with the intent to answer it correctly. These qualitative differences in examinee test-taking behavior bias parameter estimates for item response models that do not control for such behavior. This paper briefly reviews some of the existing item response models that account for differences in examinee test-taking behavior. A mixture Rasch model (MRM) with item response time components is proposed as an alternative, and it is evaluated through a simulation study and application to real test data.

Research on Test Speededness

Test speededness is the extent to which some examinees either (a) fail to answer all items before reaching the time limit, or (b) provide rapid responses to some items in order to complete the test within the time limit. The latter type of test-taking behavior is referred to as rapid-guessing behavior, and it is qualitatively distinct from solution behavior that is characterized by the intent to answer an item correctly (Schnipke & Scrams, 1997). The result of test speededness and rapid guessing is local dependence among end-of-test items (Douglas, Kim, Habing, & Gao, 1998). Oshima (1994) demonstrated that this violation of the assumption of local independence in item response theory (IRT) affects estimation of ability and item parameters, though the latter are affected to a greater extent. For the three parameter logistic model in particular, the item discrimination and item difficulty parameters tend to be over estimated, while the pseudo-guessing parameter is underestimated. Several item response models based on mixture modeling techniques (see McLachlan & Peel, 2000) have been developed to control for the influence of test speededness on parameter estimation. A mixture model, or latent class model, identifies examinees who belong to different subpopulations, such as a rapid-guessing class and a solution behavior class.

The extended HYBRID model (Boughton & Yamamoto, 2007; Yamamoto & Everson, 1997) allows a single examinee to have two different response models. An item response model such as the two parameter logistic model describes the unspeeded part of an examinee's response pattern and a multinomial distribution characterizes the speeded portion. The latter portion of the model does not take examinee proficiency or item difficulty into account in determining the probability of a correct response. As such, the model determines whether proficiency or random guessing underlies item response and applies the appropriate model to an examinee's response pattern. A simulation study indicated that the extended HYBRID model recovered item parameters more accurately than a standard IRT model in the presence of random responses due to test speededness (Yamamoto & Everson, 1997).

Bolt, Cohen, and Wollack (2002) used a two-class MRM with ordinal constraints to account for test speededness that was later extended to a multi-class MRM by Mroch, Bolt, and Wollack (2005). The ordinal constraints stipulate that item difficulty estimates are larger for the rapid-guessing class. Unlike the HYBRID model, the MRM uses the same model in each class, but with different estimates of item difficulty and examinee proficiency. Schnipke and Scrams (1997) noted that the proportion of correct responses in the rapid-guessing class were above chance levels for many items, which suggested that examinees were rapidly guessing but not necessarily randomly guessing. Estimation of proficiency in the MRM is consistent with Schnipke and Scrams (1997) observation in that it allows proficiency to be estimated in the rapid-guessing class. Conversely, the HYBRID models assumes that examinees are randomly guessing. Bolt et al. (2002) found that the MRM item parameter estimates for the nonspeeded class were comparable to estimates of the same items under conditions that lacked the influence of speededness. This finding was replicated by Mroch et al. (2005) who also observed that the MRM and HYBRID model produced similar results. In the context of multiple test forms, Wollack, Cohen, and Wells (2003) demonstrated that the MRM for test speededness improved equating and helped prevent against scale drift.

An advantage of the HYBRID and MRM models of test speededness is that they use the item response patterns alone to identify examinees engaging in rapid-guessing or solution behavior. These models are suitable for paper-and-pencil tests where no other information about the speed of examinee response is available. When additional information such as item response time is available, alternative techniques may be used to identify rapid-guessing and solution behavior examinees.

Item Response Time Research

Bimodal distributions are often indicative of latent classes that differ qualitatively on some outcome. Schnipke and Scrams (1997) noted that speededness manifested in bimodal item response time distributions for items near the end of a high-stakes computer-based test. The first mode occurred shortly after the item was presented to examinees, while the second mode occurred at a later response time. Rapid-guessing examinees were identified as those with an item response time near the first mode, whereas examinees engaging in solution behavior were those with a response time near the second mode. Schnipke and Scrams (1997) used a two-state mixture model (i.e. two class finite mixture model) of item response time to identify the different types of examinees, but no adjustment was made to estimates from an item response model as in the HYBRID model and MRM.

Wise and Kong (2005) discovered that bimodal item response time distributions also occur in the context of low-stakes testing. They too indicated that the bimodal distribution represented rapid-guessing behavior and solution behavior, but the behavior was ascribed to a different type examinee motivation. On a high-stakes test examinees are motivated to finish the test. As a result some examinees resort to rapid-guessing on end-of-test items. During a low-stakes test examinees rapidly guess to avoid expending effort on an item. Consequently, in a low-stakes test, rapid guessing may occur on any test item, not just those near the end of the test (Wise & Kong, 2005).

Wise and DeMars (2006) created the effort moderated item response model (EMIRM) as another alternative for adjusting parameter estimates for differences in test-taking behavior. Their model uses item response time to determine whether examinees are engaging in rapid-guessing or solution behavior. Like the HYBRID model, a different response model is applied to the responses from each type of examinee. Item responses due to solution behavior follow a three parameter logistic model, and item responses attributed to rapid-guessing behavior follow a constant probability. Wise and DeMars (2006) found that item difficulties estimated from the EMIRT were lower than those estimated from a standard IRT model. This result was consistent with those from applications of the HYBRID and MRM model even though information about item response patterns was not incorporated into the model. Consequently, Wise and DeMars (2006) demonstrated that incorporating item response time into an item response model was an effective way to adjust parameter estimates for the influence of test taking behavior.

The EMIRT model has the added advantage of using time to directly measure the extent to which someone is rapidly responding. However, this advantage may be offset by the fact that class membership is not estimated as part of the model. It is determined through visual inspection of item response time distribution plots and treated as known. Although Kong, Wise, and Bhola (in press) demonstrated that visual inspection of item response plots and finite mixture models of item response time classify examinees in a similar manner, overall model fit may be improved by estimating class membership rather than treating it as known.

Given that rapid-guessing and solution behavior examinees may be identified through bimodal item response time distributions and conditional dependencies among item responses, a model is proposed that makes use of both sources of information in identifying examinee test taking behavior and controlling for the influence of this behavior on parameter estimation. The model combines a finite mixture model of item response time with a MRM of item response.

A Mixture Rasch Model with Item Response Time Components

The model uses two outcome variables: item response time, t, and item response, x. Given $i = 1, \ldots, N$ examinees, $g = 1, \ldots, G$ latent classes, and $j = 1, \ldots, J$ items, the probability of observing an item response time follows a log-normal finite mixture distribution,

$$p(t|g, \lambda_{gj}, \tau_{gj}) \sim \text{Log-Normal}(\lambda_{gj}, \tau_{gj}), \tag{1}$$

where λ_{gj} and τ_{gj} are the class specific item response time mean and precision, respectively. The probability of a correct response, x, is given by the mixture Rasch model,

$$p(x|g, \theta_{gi}, b_{gj}) = \frac{exp(\theta_{gi} - b_{gj})}{1 + exp(\theta_{gi} - b_{gj})}, \tag{2}$$

where θ_{gi} is the ability of person i in class g, and β_{gj} is the class specific item difficulty. For the mixture distribution to be properly defined, the mixing proportions must meet the restrictions that $\sum_{g=1}^{G} \pi_g = 1$, and $0 \leq \pi_g \leq 1$. In addition, the b_{gj} were restricted to have a mean of zero within each class.

Assumptions

For Equation 1, item response times are assumed to be independent of each other, given latent class membership. The item response model in Equation 2 is based on assumptions of the MRM (see Rost, 1990).

In describing a model of response time and accuracy, van der Linden (2007) justified the assumption that item response and item response time are independent given ability and a personal speed parameter. A similar assumption is presumed to hold for the current model. The assumption is that item response time and item response are independent given latent class membership, $p(t, x|g) = p(t|g)p(x|g)$. Any dependency between item response time and item response is presumed to be attributed to latent class. Moreover, local dependencies in item response attributed to time or speededness are alleviated by conditioning on proficiency and latent class. Douglas et al. (1998, p. 144) noted that conditional dependencies due to test speededness could be removed if "time were conditioned on as well as theta." For the present model, conditioning on latent class and theta has the same result as conditioning on time and theta.

By assuming that item response time and item response are independent given latent class membership, the complete conditional distributions for the model described by Equation 1 and 2 remain unchanged when combined into a single model. As such, item response and item response time contribute to the estimation of latent class membership, g, and the mixing proportions, π_g, but, given latent class, item response does not contribute to estimation of λ_{gj} or τ_{gj} and item response time does not contribute to estimation of θ_{gi} or b_{gj}.

Estimation and Model Comparison

The parameters of the combined model were estimated with Markov chain Monte Carlo techniques using the program OpenBUGS 3.0.3 (Spiegelhalter, Thomas, Best, & Lunn, 2007). Starting values for the item response time means and variances and the latent class membership indicator were obtained from a mixture model analysis of item response time that used maximum likelihood estimation. This particular model was a multivariate log-normal finite mixture distribution with a diagonal variance-covariance matrix. The parameters were estimated using the R package mclust 3.1-2 (Fraley & Raftery, 2007). All other starting values were randomly generated.

The prior distributions were

$$
\begin{aligned}
b_{gj} &\sim \text{Normal}(0, 1) \\
\theta_{gi} &\sim \text{Normal}(\mu_g, 1) \\
\mu_g &\sim \text{Normal}(0, 1) \\
\lambda_{1j} &\sim \text{Normal}(0, 100) \\
\lambda_{2j} &\sim \lambda_{1j} + \text{Normal}(0, 100)\text{I}(0,) \\
\tau_{gj} &\sim \text{Gamma}(.01, .01) \\
\sigma_{gj} &= 1/\sqrt{\tau_{gj}} \\
(\pi_1, \pi_2) &\sim \text{Dirichlet}(.5, .5).
\end{aligned}
$$

Given the proposed interpretation of the latent classes, the item response time means for class two, the solution behavior class, were constrained to be higher than those for class one, the rapid-guessing class. This was accomplished by restricting the second class's prior for mean item response time to a value that was larger than the first class's prior for mean item response time. The I(0,) function in the prior for λ_{2j} stipulates that only positive draws from a normal distribution were permitted.

The likelihood for the proposed model is

$$L(\mathbf{x}, \mathbf{t}) = \prod_{i=1}^{N} \prod_{j=1}^{J} p(x_{ij}) q(x_{ij}) p(t_{ij}), \tag{3}$$

where $p(x_j) = \sum_{g=1}^{G} \pi_g p(x_j | g, \theta_{gi}, b_{gj})$, $q(x_{ij}) = \sum_{g=1}^{G} \pi_g [1 - p(x_j | g, \theta_{gi}, b_{gj})]$, and $p(t_{ij}) = p(t_i | g, \lambda_{gj}, \tau_{gj})$. One- and two-class solutions for the extant data were compared with the Bayesian information criterion,

$$\mathrm{BIC} = -2\log L(\mathbf{x}, \mathbf{t}) + m \log N, \tag{4}$$

where m is the number of free model parameters and $L(\mathbf{x}, \mathbf{t})$ is given by Equation 3.

Variations

Although the MRM with item response time components is presented as a single model, it is flexible enough to be adapted to specific applications. The only constraints used in this model are those imposed on the item response time means. Additional constraints may be used to tailor the model for specific situations. For example, for items at the beginning of a test, the item difficulty and average item response time values for the rapid-guessing class may be set equal to those for the solution behavior class. Like the MRM for test speededness, these parameters could only be free to vary between classes for end-of-test items. Similar equality constraints among item response time means could be imposed on items that do not appear to have a bimodal distribution. By imposing constraints, the model can account for a variety of situations.

Method

Simulation Study

Parameters for the simulation study were obtained from a preliminary two-class analysis of the extant data described below. Proficiency values were drawn from a N(-0.56, 1) distribution and a N(0.78, 1) distribution for latent class one and two, respectively. Note that N(μ, σ) denotes a normal distribution with mean μ and standard deviation σ. Rapid-guessing examinees typically represent only a small portion of the tested population. Therefore, the number of examinees in the first latent class was fixed to 15% of the total sample size. Twenty-five item parameter estimates from the extant data analysis served as the generating item parameters after rescaling. Given a simulee's latent class membership, proficiency value, and class specific item difficulty for an item, the simulee's response vector was created by drawing a value from a Bernoulli($p(x|g, \theta_{gi}, b_{gj})$) distribution for each j. In a similar manner, class specific item response time was generated for each item from a log-normal distribution with mean and variance set to the values estimated from the same 25 items in the extant data. Ten replications of 500 simulees and ten replications of 2,000 simulees were created.

Convergence was checked after each run by evaluating autocorrelation and time series plots as well as the Brooks-Gelman multivariate potential scale reduction factor (MPSRF; Brooks & Gelman, 1998). Sinharay (2004) recommended the use of the MPSRF due to its apparent sensitivity to lack of convergence.

Extant Test Data

The extant data were collected from a spring 2004 administration of the Information Literacy Test (ILT). The ILT was administered by computer to a random sample of $N = 524$ college sophomores. Cronbach's alpha was 0.88 for this administration of the test. Additional psychometric characteristics are described by Cameron, Wise, and Lottridge (2007). A one- and two-class solution was applied to the data and the two models were compared using the BIC (Equation 4). If the BIC indicates that the one-class solution fits better than the two-class solution, the effect of examinee test taking behavior could be considered negligible. Conversely, if evidence suggests that the two-class solution fits better, the influence of examinee test taking behavior must be accounted for by two latent classes. Otherwise, parameter estimates would be biased in the manner described by Oshima (1994) and Douglas et al. (1998).

Results

Simulation Study

Preliminary analysis of simulated data suggested that adequate convergence was reached by running four chains for 15,000 iterations past a burn-in of 5,000 iterations and thinning to every fifth iteration. The analysis of a single data set took one to two weeks to complete on a Pentium IV, 3.19 GHz machine. After each analysis, the MPSRF was computed. All MPSRF values were either 1.03 or 1.04 in the $N = 500$ and $N = 2,000$ condition, indicating that the chains had converged in each analysis.

The mixing proportion, π_g, was estimated rather well in all conditions and as indicated by the small RMSE values (see Table 1). This result is likely due to the amount of information brought to bear on estimating class membership and the mixing proportion. Both item response and item response time contribute to estimation of these parameters. Although the mixing proportion appears to be estimated equally well in both classes, the mixing proportion was only freely estimated for one class. The other class's mixing proportion was constrained to be one minus the other class's mixing proportion.

Given that class two represented a larger proportion of simulees, the parameters for class two were estimated more accurately than those for class one (see Table 1). The RMSE values were notably larger for the $N = 500$ condition, particularly among the class one parameter estimates. Not only was the total sample size smaller in the $N = 500$ condition, but class one comprises as few as 75 examinees. The small size of this class likely contributed to the large RMSE values for class one.

Analysis of ILT

Figure 1 shows kernel density plots (see Silverman, 1986) of log response times for six ILT items that were typical of all ILT items. The bimodal nature of these densities is evident. The first mode appears early and its density is not very large, suggesting that only a few examinees, the rapid-guessers, are in this part of the distribution. The second mode appears later and its density is much lager indicating that more examinees, the solution behavior class, are in this region. Although most ILT item response time distributions appeared to be bimodal, a few showed signs of a third mode at an even later time than the second mode, but it did not seem distinct enough to account for it in the model.

Four chains of 70,000 iterations were run, and the first 39,999 iterations were discarded as burn-in. Each chain was thinned to every fifth iteration. The MPSRF for the two-class solution was 1.05. This value is close to one and suggests that the precision of estimation may not be improved much by running the chains for more iterations. The two-class MRM with item response time components indicated that 15% of examinees were in the rapid-guessing class and the remaining 85% were in the solution behavior class. The average proficiency of the rapid guessing class was also notably lower ($\mu_1 = -0.19$) than the average proficiency for the solution behavior class ($\mu_2 = 1.42$). Item response times were more variable for the rapid-guessing class than the solution behavior class. The average σ_{gj} was 1.03 and 0.48 for the rapid-guessing and solution behavior classes, respectively (details not shown).

Figure 2 shows the standardized item difficulty and log item response time mean estimates for Class 1, the rapid-guessing class, and Class 2, the solution behavior class. Interestingly, most of the item parameter estimates are not ordered for the two classes. Only 45% of the standardized item difficulty estimates were larger for the rapid-guessing class. Moreover, the two largest estimates were found in the solution behavior class. This pattern does not seem to agree with the ordinal constraints of the Bolt et al. (2002) MRM for test speededness. However, with ordinal constraints on the item difficulties, quantitative differences between examinees in each class may not only produce a shift in mean proficiency but also a shift in item difficulty with further differences among item parameter estimates reflecting qualitative difference between classes. In the MRM with item response time components, ordinal constraints are not imposed on the item difficulties. Therefore, quantitative difference between examinees are manifest only in the mean proficiency values for each class.

The items plotted in Figure 1 are illustrated again in Figure 3. The dashed lines in Figure 3 correspond to the estimated finite mixture components of item response time. The fit of the components appears to be rather close to the observed distributions. Indeed, the BIC for the entire model was 273,527.07 for the one-class solution, but it was 256,660.40 for the two-class solution. The two-class solution resulted in better model fit.

Discussion

Rapid guessing and solution behavior manifest in bimodal item response time distributions and conditional dependencies among item responses. Existing models make use of one type of information, either item response or item response time, but not both. A MRM with item response time components was proposed to take advantage of both sources of information that reflect differences in test taking behavior. A simulation study demonstrated that parameters may be recovered well. The mixing proportion, in particular, was very close to the nominal value. This result was likely due to using both item response and item response time in its estimation. The other trends in the simulation results were expected. The accuracy of the estimates improved as the group size and total sample size increased.

Analysis of the ILT indicated that a two-class solution fit better than a one class solution. This finding supported the use of a latent class approach to accounting for differences in test taking behavior. Another interesting result was that the item difficulty estimates were not ordered for the two classes. That is, the rapid-guessing class did not have larger item difficulty estimates than the solution behavior class. It is likely that all quantitative differences between the latent classes were reflected in the class mean proficiency. Imposing ordinal constraints on the item difficulty estimates as in the MRM of test speededness, would likely reduce the difference between class means while increasing the differences between item parameter estimates.

Differences in test taking behavior result in qualitatively distinct item response patterns and bimodal item response time distributions. If not taken into account, these behavioral results have an adverse impact on parameter estimation. When item response and item response time is available, the MRM with item response time components can accommodate difference in test taking behavior and accurately estimate item difficulty parameters.

References

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.

Boughton, K. A., & Yamamoto, K. (2007). A hybrid model for test speededness. In M. von Davier & C. H. Carstensen (Eds.), *Multivaraite and mixture distribution Rasch models.* New York: Springer.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434-455.

Cameron, L., Wise, S. L., & Lottridge, S. M. (2007). The development and validation of the information literacy test. *College and Research Libraries*, *68*, 229-236.

Douglas, J., Kim, H. R., Habing, B., & Gao, F. (1998). Investigating local dependence with conditional covariance functions. *Journal of Educational and Behavioral Statistics*, *23*, 129-151.

Fraley, C., & Raftery, A. (2007). *The mclust package.* (http://www.stat.washington.edu/mclust/)

Kong, X. J., Wise, S. L., & Bhola, D. S. (in press). Setting the response time threshold parameter to differentiate solution behavior from rapid-guessing behavior. *Educational and Psychological Measurement*.

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York: John Wiley & Sons.

Mroch, A. A., Bolt, D. M., & Wollack, J. A. (2005, April). A new multi-class mixture rasch model for test speededness. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal, Quebec.

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.

Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, *34*, 213-232.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* New York: Chapman and Hall/CRC.

Sinharay, S. (2004). Experiences with markov chain monte carlo convergence assessment. *Journal of Educational and Behavioral Statistics*, *29*, 461-488.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2007). *Openbugs: User manual, version 3.0.2.* (Retrieved from http://mathstat.helsinki.fi/openbugs/)

van der Linden, W. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*, 287-308.

Wise, S. L., & DeMars, C. E. (2006). An application of item response time: The effort moderated irt model. *Journal of Educational Measurement*, *43*, 19-38.

Wise, S. L., & Kong, X. J. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, *16*, 163-183.

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., & Everson, H. (1997). Modeling the effects of test length and test time on parameter estimation using the hybrid model. In J. Rost (Ed.), *Applications of latent trait and latent class models in the social sciences* (p. 89-98). Münster, Germany: Waxmann.

Appendix

OpenBUGS Code for the MRM with Item Response Time Components

```
model{
for(i in 1:N){
    for(j in 1:NI){
        rt[i,j] ~ dlnorm( lambda[class[i], j], tau[class[i], j])
        p[i,j] <- exp(theta[i]-b[class[i], j])/(1+exp(theta[i]-b[class[i], j]))
        x[i,j] ~ dbern(p[i,j])
    }
}
for(i in 1:N){
    class[i] ~ dcat(pi[])
    theta[i] ~ dnorm(mup[ class[i] ], 1)
}
for(j in 1:NI){
    beta[1, j] ~ dnorm(0,1)
    beta[2, j] ~ dnorm(0,1)
    lambda[1, j]~ dnorm(0, 1.0E-2)
    lambda[2, j]<-lambda[1, j]+shift[j]
    shift[j] ~ dnorm(0, 1.0E-2)I(0,)
    for(g in 1:G){
        tau[g, j] ~ dgamma(0.01, 0.01)
        sigma[g, j]<-1/sqrt(tau[g, j])
    }
}
pi[1:G] ~ ddirch(alpha[])
```

```
for(g in 1:G){
    mup[g] ~ dnorm(0,1)
}
for(j in 1:NI){
    b[1, j]<-beta[1, j]-mean(beta[1, 1:NI])
    b[2, j]<-beta[2, j]-mean(beta[2, 1:NI])
}
for(i in 1:N){
    for(j in 1:NI){
        sd[i,j]<-sigma[class[i], j]
        var[i,j]<-pow(sd[i,j], 2)
        const[i,j]<-1/(rt[i,j]*sd[i,j]*sqrt(2*3.141592654))
        lrt[i,j]<-log(rt[i,j])
        eSqrd[i,j]<-pow(lrt[i,j]-lambda[class[i],j], 2)
        pRt[i,j]<-const[i,j]*exp(-eSqrd[i,j]/(2*var[i,j]))
        logLikeTemp[i,j] <- x[i,j]*log(p[i,j]) + (1-x[i,j])*log(1-p[i,j]) + log(pRt[i,j])
    }
}
logLike <-sum(logLikeTemp[1:N, 1:NI])
}
```

Table 1

RMSE and Bias for Simulated Data

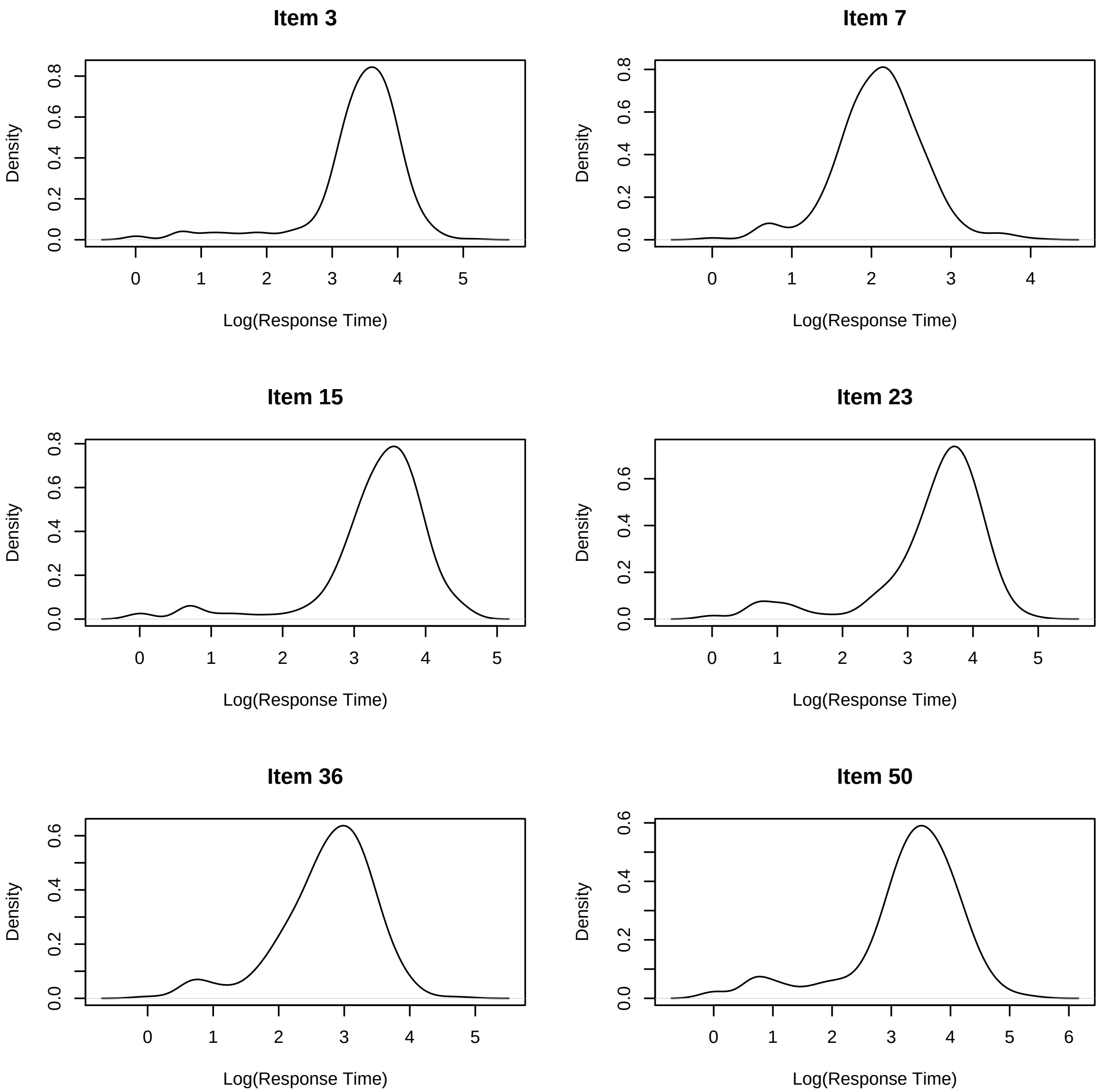
		Class One		Class Two	
N	Parameter	RMSE	Bias	RMSE	Bias
500	b	0.2770	0.0032	0.1752	0.0000
	μ_g	0.2901	0.2353	0.0302	0.0126
	λ	0.1263	-0.0139	0.0347	0.0015
	σ	0.0888	0.0069	0.0211	0.0000
	π	0.0016	0.0016	0.0016	-0.0016
2,000	b	0.1259	0.0000	0.0751	0.0000
	μ_g	0.0378	0.0159	0.0727	0.0708
	λ	0.0648	0.0030	0.0162	-0.0003
	σ	0.0404	-0.0030	0.0133	-0.0001
	π	0.0006	0.0005	0.0006	-0.0005

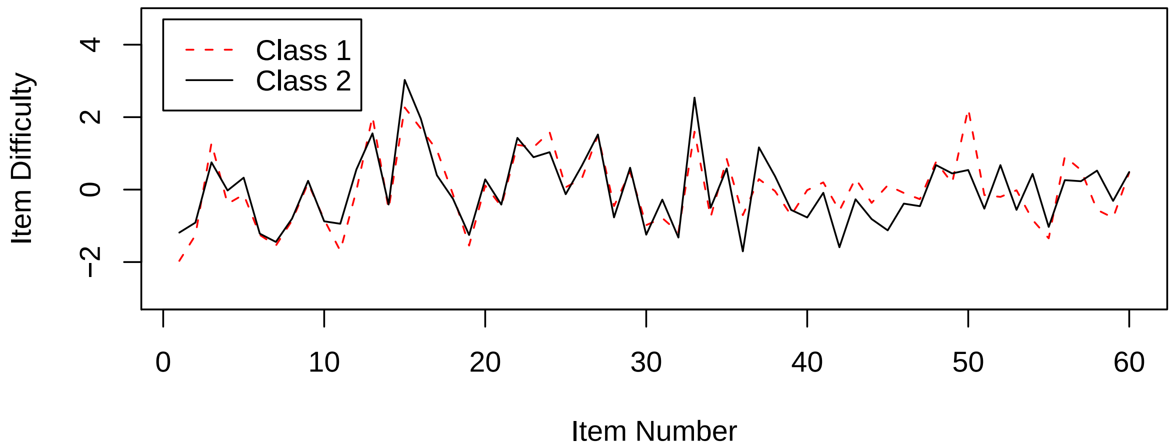
Figure Captions

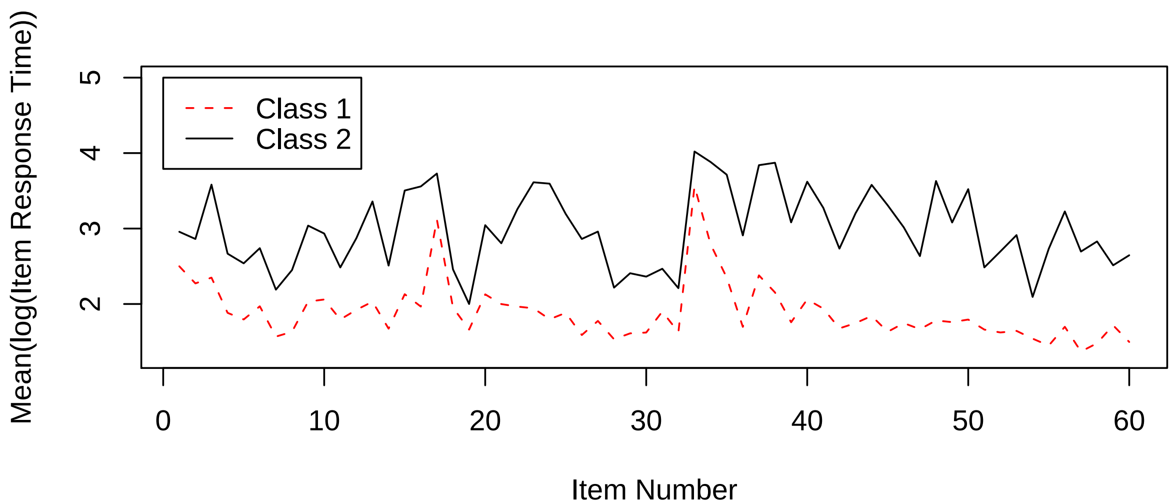
Figure 1. Kernel Density Plots of Selected Item Response Time Distributions

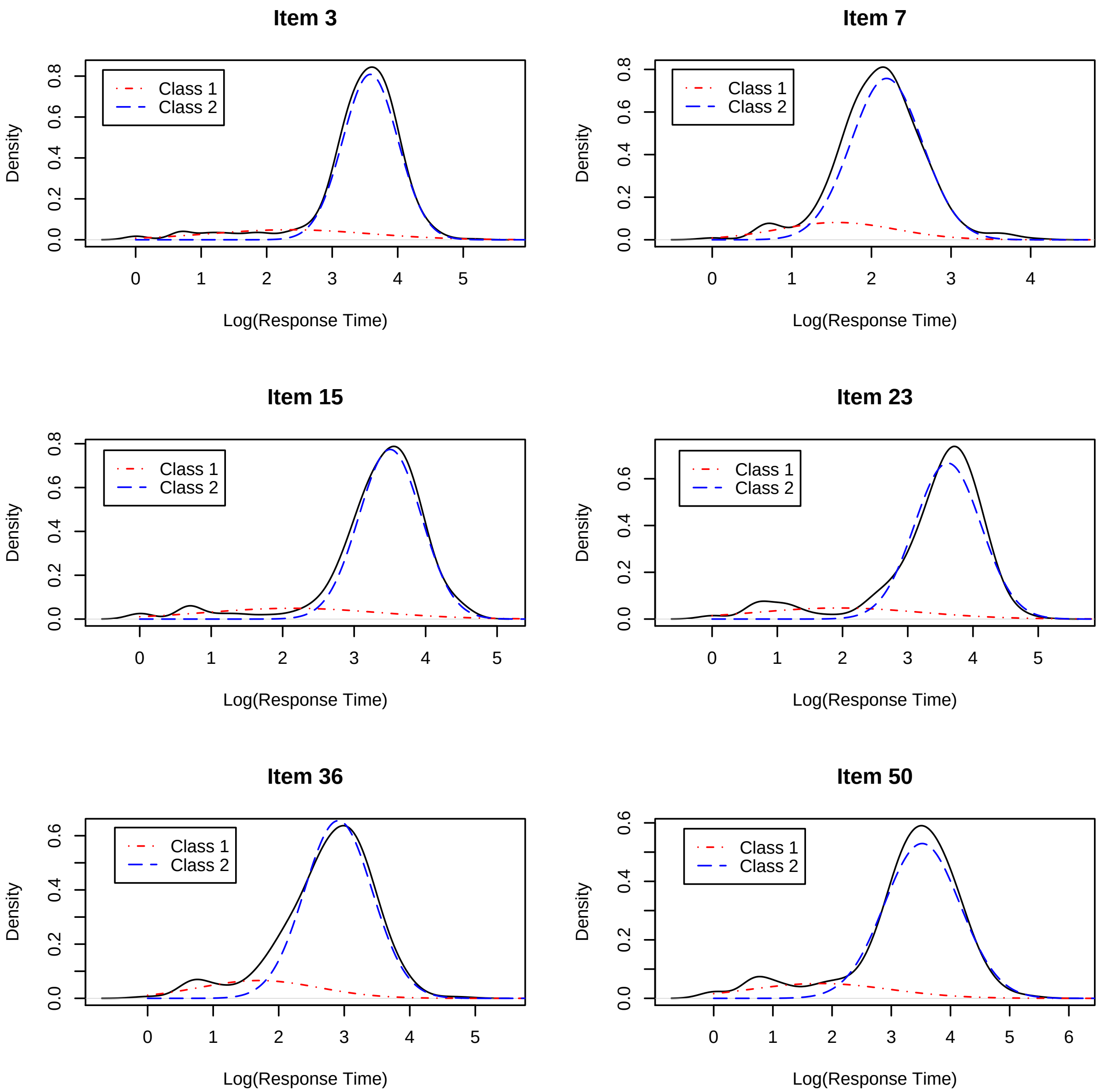
Figure 2. Standardized Item Difficulty and Log Item Response Time Estimates for Class 1 (rapid-guessing) and Class 2 (solution behavior)

Figure 3. Kernel Density Plots with Fitted Log-Normal Densities


Item 3
Item 7
Item 15
Item 23
Item 36
Item 50
Density
Log(Response Time)


Class 1
Class 2
Item Difficulty
4
2
0
−2
0
10
20
30
40
50
60
Item Number


Class 1
Class 2
Mean(log(Item Response Time))
5
4
3
2
0
10
20
30
40
50
60
Item Number


Item 3
Item 7
Item 15
Item 23
Item 36
Item 50
Class 1
Class 2
Density
Log(Response Time)