ED 454 268 TM 032 877

AUTHOR Matthews-Lopez, Joy L.; Hombo, Catherine M.
TITLE Modeling the Hyperdistribution of Item Parameters To Improve the Accuracy of Recovery in Estimation Procedures.
SPONS AGENCY Educational Testing Service, Princeton, NJ.
PUB DATE 2001-00-00
NOTE 25p.; Supported by Educational Testing Service Research Allocation Project 862.07.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Estimation (Mathematics); Monte Carlo Methods; Statistical Distributions; *Test Construction
IDENTIFIERS *Item Parameters

ABSTRACT

The purpose of this study was to examine the recovery of item parameters in simulated Automatic Item Generation (AIG) conditions, using Markov chain Monte Carlo (MCMC) estimation methods to attempt to recover the generating distributions. To do this, variability in item and ability parameters was manipulated. Realistic AIG conditions were simulated, and the SCORIGHT computer program was used to estimate item parameters and simulee ability. There were indications that the MCMC estimation failed to converge in the 2000 cycle run. Histograms for some of the items show that the MCMC procedure had not yet converged for the individual runs or that the program was not operating correctly, and that the former was more likely. It was uncertain that valid inferences would be made based on the analyses. Follow-up work is planned, using 25,000 iterations. (SLD)

Modeling the Hyperdistribution of Item Parameters to Improve the Accuracy of Recovery in Estimation Procedures

Joy L. Matthews-López

Catherine M. Hombo

Educational Testing Service

The work reported herein was supported by Educational Testing Service Research Allocation Project 862.07. The opinions expressed herein are solely those of the authors and do not necessarily represent those of Educational Testing Service. The authors would like to express their thanks to Jane Rogers for her work on the data simulations.

Introduction

Large-scale testing programs expend tremendous amounts of fiscal and human resources on item development, item pool maintenance, and item security. Item quality is a function of several variables, the greatest of which is quality item authoring. Regardless of testing platform, items must be written to fit specific guidelines and constraints as dictated by tables of specifications. After an item has been initially written, the journey to becoming an operational item can be long and arduous.

All newly written items are subject to reviews for fairness and for written quality. After fairness review, items are prepared for piloting and preliminary calibration. Prime target locations for piloting are carefully identified and schools and/or institutions are contacted for participation. Finding such schools is becoming an increasingly difficult problem due to the already heavy testing schedule of most institutions. Regardless of incentives to participate, schools have limited time to allocate to non-essential testing. It is only after pre-testing, preliminary item analyses, and item calibration that an item is approved for operational use. Though item writers may be well trained and experienced, it remains virtually impossible for even the best of writers to consistently construct items to specific item parameters.

Computer-based testing programs place additional requirements, beyond the traditional test assembly constraints, upon their items. In order to sustain the validity of any computer-based testing program, special care must be given to the maintenance of the item pool from which its items are selected. Many times this translates into an increased number of specifically constrained items. The problem arises when items of

specific difficulty and/or discrimination are needed, in combinations of varied content constraints, to fill the requirements of target test information functions.

One possible response to the challenge of meeting the demand for high quality, parameter-specific, cost efficient items is automatic item generation (AIG) (Bejar, 1993; Kyllonen, (in press); Meisner, Luecht, & Reckase, 1993). Through this process of item generation it is expected that items of specific psychometric qualities can reliably be produced. If AIG is shown to be a viable option for item creation, then the demands on human item writers can be reduced to a more manageable level.

Automatic Item Generation (AIG) is defined, for the purpose of this paper, as a process used to create groups of items. This process, usually performed by a computer, consists of the creation of item "shells". These shells are then used, by the computer, to generate an unlimited number of "family" items, sometimes referred to as isomorphs. The relationship between items, known as isomorphism, is a very strong assumption. Isomorphism implies that variability among the item parameters of a family is negligible. That is to say that any two within-family isomorphs are assumed to have similar item characteristics, properties, and most specifically, item parameters.

Item shells consist of both variable and fixed parts, as determined by the designer. Variable parts are usually constrained in order to control the range of item difficulty. This is done in an attempt to support the assumption of isomorphism. As the term implies, fixed parts of a shell are non-varying and as such, carry equally into each of the offspring family isomorphs. Without familial isomorphism, AIG will not function as desired. It is therefore crucial to the success of future AIG endeavors to carefully examine the assumption of familial isomorphism.

Current AIG projects are underway. Specific areas of research include, but are not limited to, spatial reasoning, GRE quantitative reasoning, mathematics items, survey items, and nonverbal ability items. In light of the awesome potential of this approach to item production, and of the enormous implications to fiscal and human resources, ensuring the stability of the assumptions underpinning this procedure is well warranted.

Purpose of the Study

The purpose of this study is to examine the recovery of item parameters in simulated AIG conditions, using Markov chain Monte Carlo (MCMC) estimation methods to attempt to recover the generating distributions. To do this, we will manipulate variability in item and ability parameters. It should be noted that in any simulation work, two types of bias exist: intentional and unintentional. Intentional bias is that which is controlled in the research design. Unintentional bias, on the other hand, is the result of estimation procedures. It is virtually impossible to accurately partition total bias into these two subparts. It is however, possible to compute the difference between generating values and estimated values, as calculated by SCORIGHT (Wang, Bradlow, & Wainer, 2000a), a new IRT estimation program that uses MCMC methods.

Study Design

This paper is one phase of a larger research project that is designed to examine the AIG assumption of familial isomorphism and the impact of using AIG in operational situations. In order to simulate realistic AIG conditions, the following procedures were followed in simulating the data:

(1) A set of generating parameters was selected. Items were generated such that the test had 25% of its items with difficulty below $b = -1.5$, 25% were above

b = 1.5, and the remainder were contained in the closed interval [-1.5, 1.5]. Item discrimination parameters (*a*) were drawn such that .7 was the minimum value. This assured that all item discriminations sampled from the distributions were positive.

(2) Item parameters for each simulee were drawn from a specified distribution. For the *a*-variance=0 and *b*-variance=0 condition (denoted herein as a0b0), each simulee received the exact same set of item parameters. For all other conditions, at least one item parameter, and possibly both, were randomly drawn for the specified distribution for each examinee. For these cases, each simulee potentially received a different set of items.

(3) Item difficulty parameters were sampled from a normal distribution with the mean set at the generating value and the variance set according to the condition. The item slope parameters were sampled from a lognormal distribution with the mean set at the generating value and the variance set according to the condition. For both parameters, the variance conditions were $\sigma^2=\{0.0, 0.3\}$.

(4) A fully crossed design resulted in four variance-combination conditions a0b0, a0b3, a3b0, and a3b3, where the number next to the parameter indicates the level of variance in the parameter hyperdistribution multiplied by 10.

(5) The original full data set consisted of 10 replications, each consisting of response date from 5000 simulees to 50 items.

(6) The ability distribution from which each simulee's true ability was sampled was defined to be *N*(0, 1).

(7) All examinee responses were simulated as dichotomously scored. All examinees responded to all items.

(8) Response data for each simulee was drawn to conform to the 3-parameter logistic (3PL) IRT model, using the generating ability (θ) for all replications.

(9) SCORIGHT, a new IRT-based scoring and parameter estimation computer program (Wang, Bradlow, & Wainer, 2000a), was used to estimate item parameters and simulee ability.

The SCORIGHT computer program functions within a fully Bayesian framework which uses Markov chain Monte Carlo procedures. SCORIGHT uses Gibbs sampling methods for inference. In order for the inferences to be valid, the Gibbs sampler must "converge" (Wang, Bradlow, & Wainer, 2000b). To increase the likelihood of convergence, a reasonable number of iterations must be allowed for "burn-in". In this study, 2000 iterations were run, allowing for 1000 extractions after convergence, following the example provided in the manual (Wang et al, 2000b). The minimum number of iterations needed for convergence is unknown, for convergence depends on the data as well as initial parameter values (Wang, Bradlow, & Wainer, in press).

SCORIGHT is designed to accurately estimate ability and item parameters for tests composed of discrete items or groups of items connected by something ("testlets", which are a group of items thought to violate the IRT assumption of local independence). When a user indicates that the test contains testlets, SCORIGHT is designed to assess the degree of local dependence and makes adjustments to the estimates, accordingly. In this study, all items are generated to be discrete, locally independent items.

Response data for each simulee was drawn to conform to the 3-parameter logistic (3PL) IRT model, as defined by the following function:

$$P_i(\theta_n) = c_i + (1 - c_i)\frac{e^{Da_i(\theta_n - b_i)}}{1 + e^{Da_i(\theta_n - b_i)}}$$

where n indexes examinees, i indexes items, and c_i is the pseudo-guessing parameter. For information relating to the model used with polytomous items, the reader is referred to the user's guide for SCORIGHT (Wang et al, in press).

The results of the SCORIGHT analysis were used to evaluate performance over generations. Since the data in this study was simulated, the "true" values for all parameters and their generating distributions were known. This allowed for evaluation of SCORIGHT's performance in parameter recovery, both in terms of the point estimates and the posterior distributions estimated by the program.

In the fully crossed design of this study, four conditions were examined: a0b0, a0b3, a3b0, and a3b3, where the number next to the parameter indicates the level of variance in the parameter distribution multiplied by 10. The burn-in for SCORIGHT was the first 1000 out of a total of 2000 cycles. 5000 simulees were used to ensure adequate sample size. Each test consisted of 50 items so as to model a realistic testing situation with reasonable internal consistency. Condition a0b0 was used as a base-line measure for this study.

Methods

Posterior distributions of the estimated parameters were obtained via SCORIGHT's MCMC estimation procedure. For each data set, SCORIGHT was run 2000 cycles, with the final 1000 draws retained. Histograms of these final 1000 draws were produced to facilitate comparison of the resulting posterior distributions for

difficulty and discrimination parameters to the generating distributions. Plots of moving averages were also computed, in order to evaluate the convergence of the analysis runs.

Results

Unfortunately, there are indications that the MCMC estimation failed to converge in the 2000 cycle runs. Based on the guidance provided in the SCORIGHT manual (Wang, Bradlow, & Wainer, 2000b), it was believed that 2000 iterations would be sufficient to obtain convergence (assumed to occur within the first 1000 draws from the posterior distribution). Extracting all draws past convergence, each run was based upon 1000 data points.

Histograms of the posterior distributions of *a* and *b*-parameters were examined under each of the four design conditions a0b0, a0b3, a3b0, and a3b3 for a representative subset of items. Items #1, #18, #35, and #50 were selected, as they represent the range of item difficulty in the test. Although all items were examined for anomalies, due to space limitations, only these four items were selected for reporting.

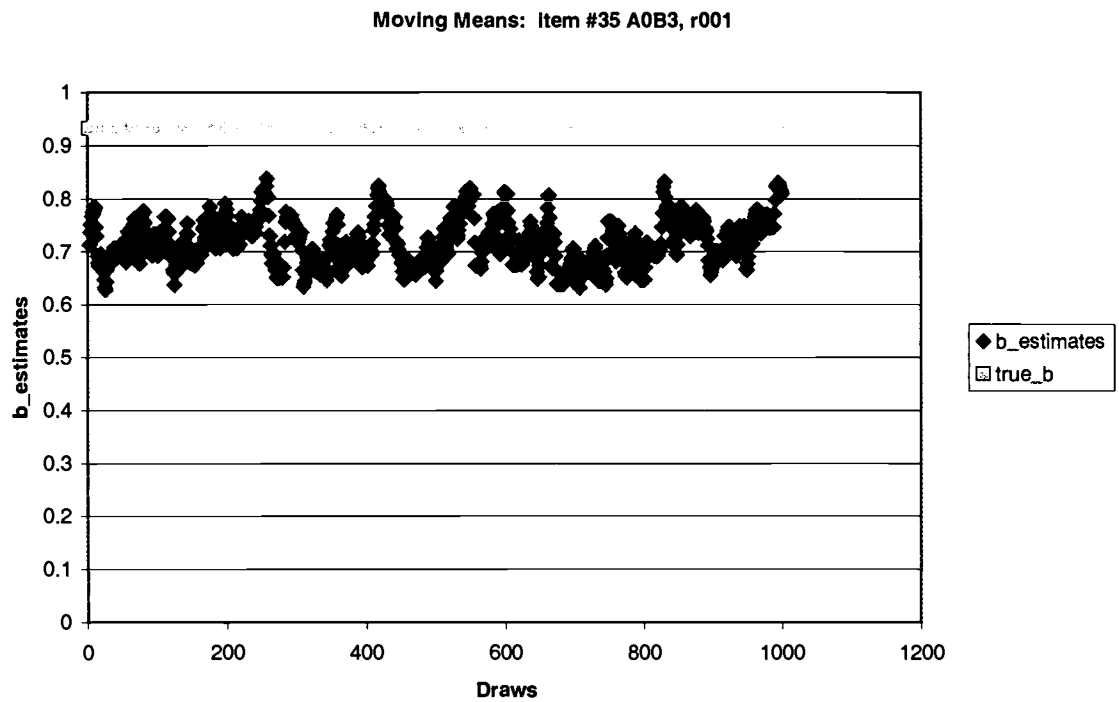
"True values" are known since data for this project was generated via pre-specified guideline and are indicated in the following table.

Insert Table 1 about here

As mentioned earlier in this paper, valid inferences are dependent on convergence. Consider the following graphs for item #1.

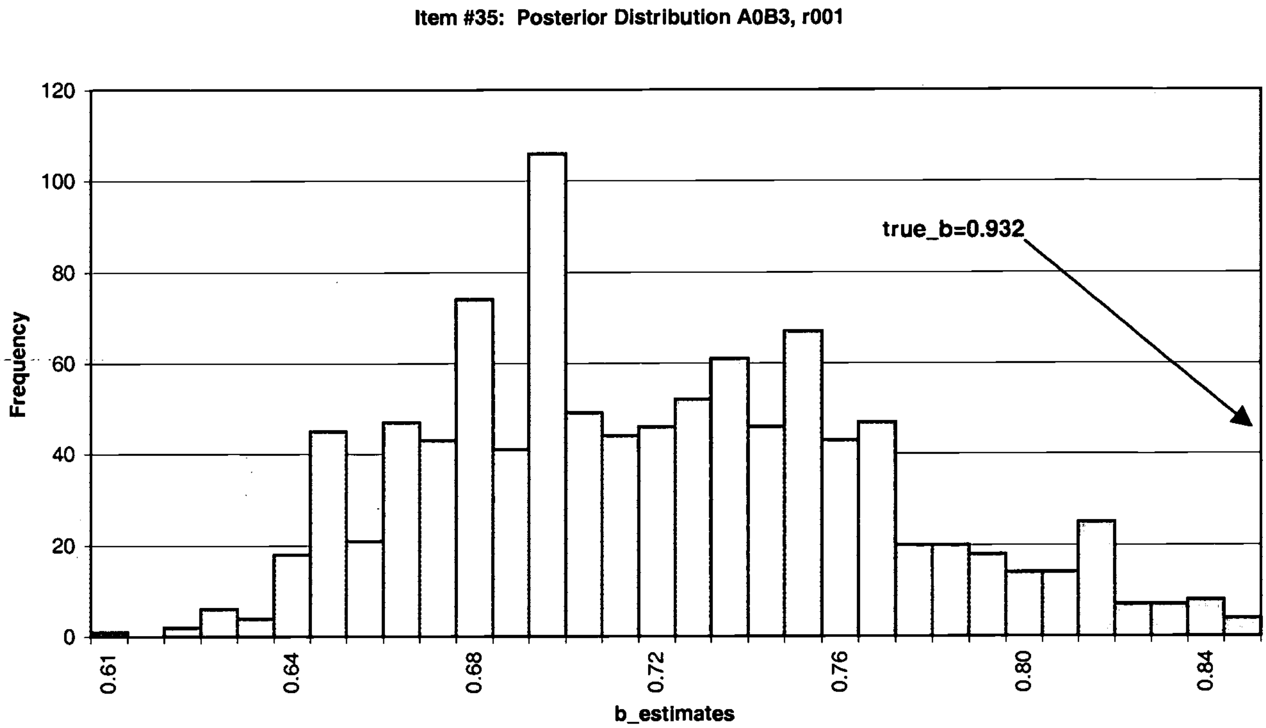
In the first chart, a moving average has been plotted. From this plot it is apparent that convergence failed to occur since the *b*-estimates failed to locate the true value of *b* (in the plot as the horizontal line).

(Chart 1)



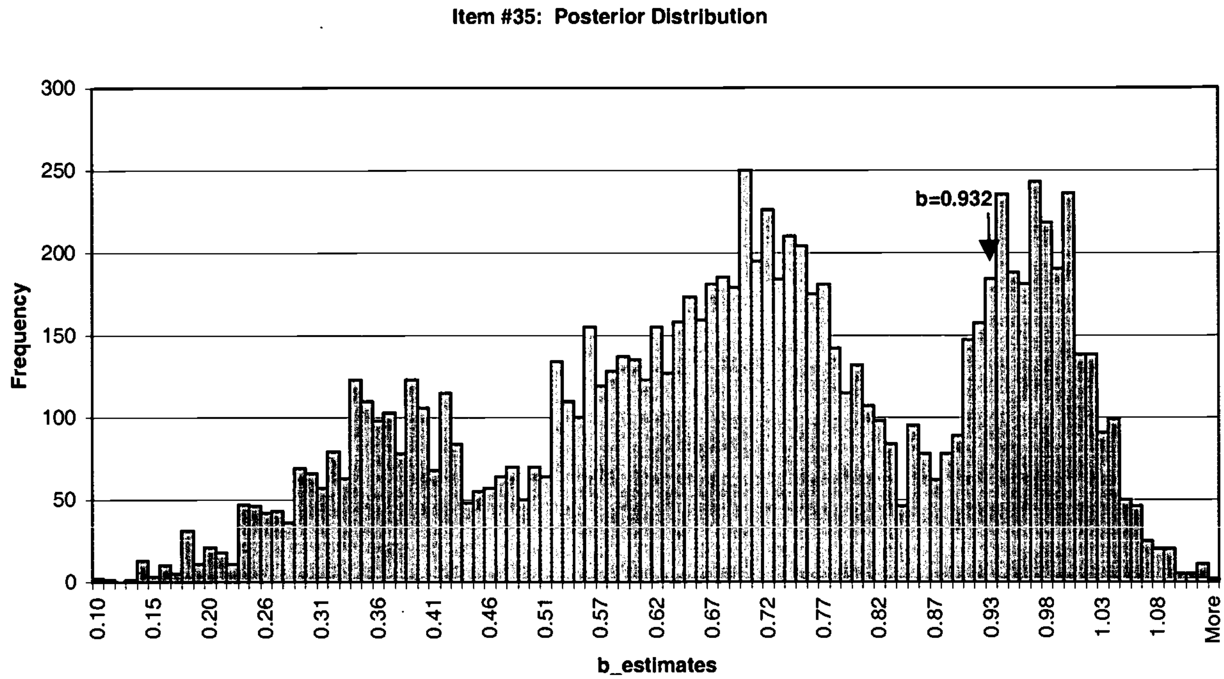
In the second chart, a histogram has been plotted. Again it is apparent that convergence has failed to occur since the true value of *b* isn't even included in the histogram.

(Chart 2)



The next chart contains a histogram produced from an aggregation of 10 runs of Scoright. Indicated in the histogram is the posterior distribution of the *b*-estimates. The generating value is indicated at the arrow. The reader should notice that this distribution is multimodal, yet each run of the data was generated from a normal distribution and had the same initializing value. This indicates one of two possibilities: the MCMC estimation procedure had not yet converged for the individual runs or the program was not operating correctly.

(Chart 3)



It can be deduced from the multimodal posterior distribution of the *b*-estimates that individual runs failed to converge. Initially it was thought that aggregation across runs would result in an increased clarity of interpretation. However it was instead found that when individual runs of Scoright fail to converge, as in this case, then aggregation serves no useful purpose.

When looking at the remaining three conditions, it was noted that induced variance conditions did appear to be different than baseline, zero variance conditions. However due to our lack of convergence, extreme caution needs to be exerted in any type of interpretation of the nonzero variance conditions a0b3, a3b0, and a3b3.

When examining the remaining three items, it became clear that under conditions a3b0 and a3b3, *a*-estimates behaved poorly and under conditions a0b3 and a3b3, *b*-

estimates behaved poorly. After consideration of the pooled information obtained from a variety of histograms of posterior distributions and moving means, it is uncertain that valid inferences can be made based on results of this analysis.

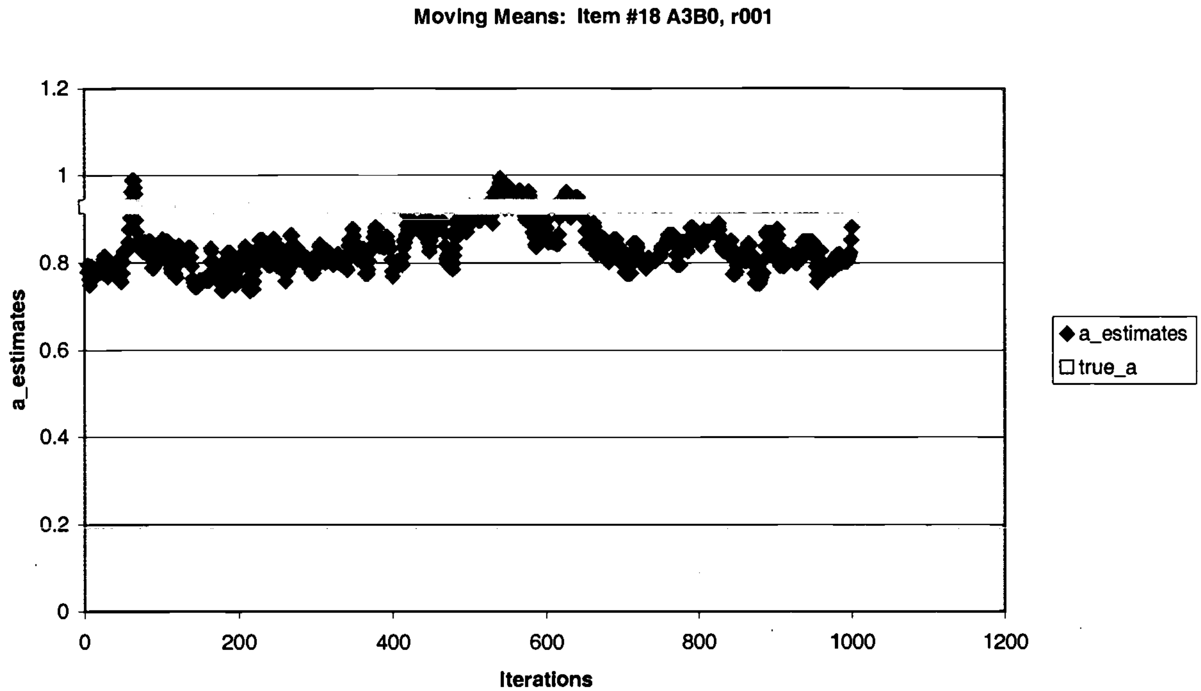
Discussion

Additional analysis of these data sets has already begun. Data is currently being re-run at 25,000 iterations, retaining the last 1,000 iterations for analysis purposes. This follow-up work is being undertaken to confirm the suspicion that convergence in the first set of analysis runs was never attained. It is important to establish these results, because if the new runs of SCORIGHT at 25,000 cycles still results in multimodal posterior distributions and non-convergent moving mean plots, then SCORIGHT is not functioning as expected. Since this is simulation data, the generating posterior distributions are known and should be recovered with reasonable accuracy at 25,000 cycles.

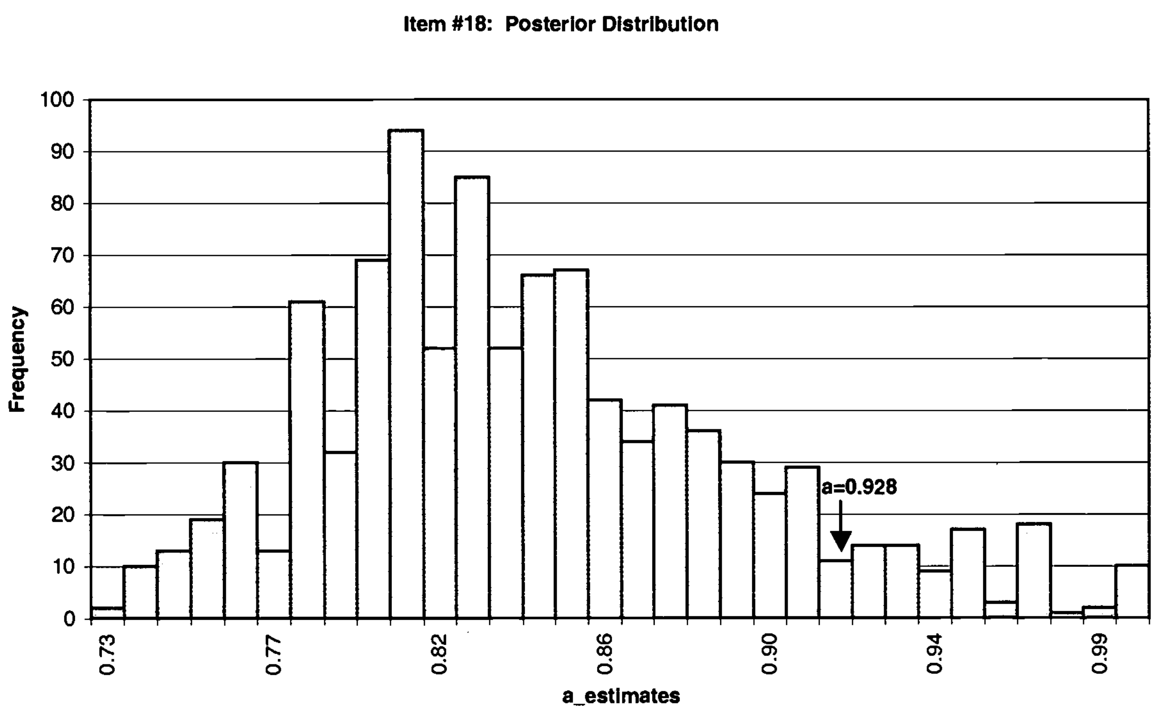
In addition to item difficulty, item discrimination and ability estimates were investigated in this project. Posterior distributions of *a* were examined and were found to be as inconclusive as with *b*. This was anticipated, given the suspicion of non-convergence of the analysis runs.

Posterior distribution and moving mean plots were constructed for item discrimination parameters. These plots are located in charts 4 and 5.

(Chart 4)



(Chart 5)

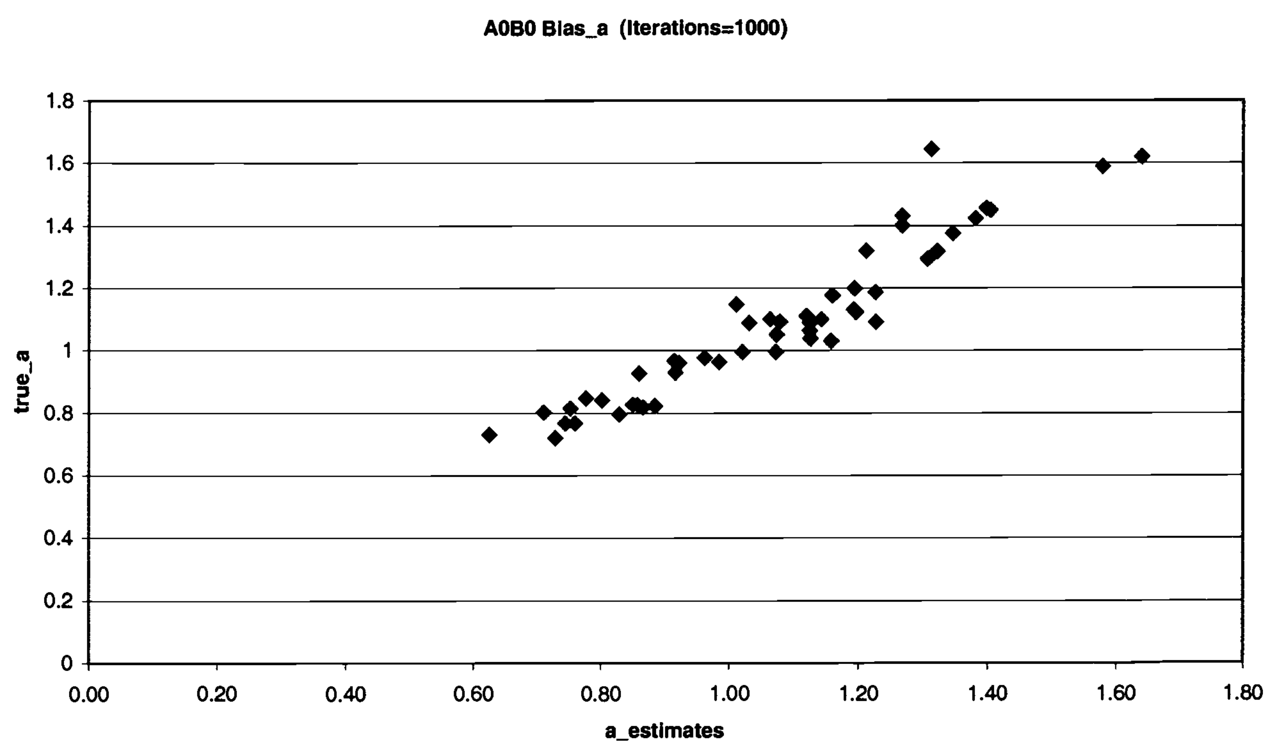


Similar plots were produced for ability distributions. The results were consistent with the theory of nonconvergence.

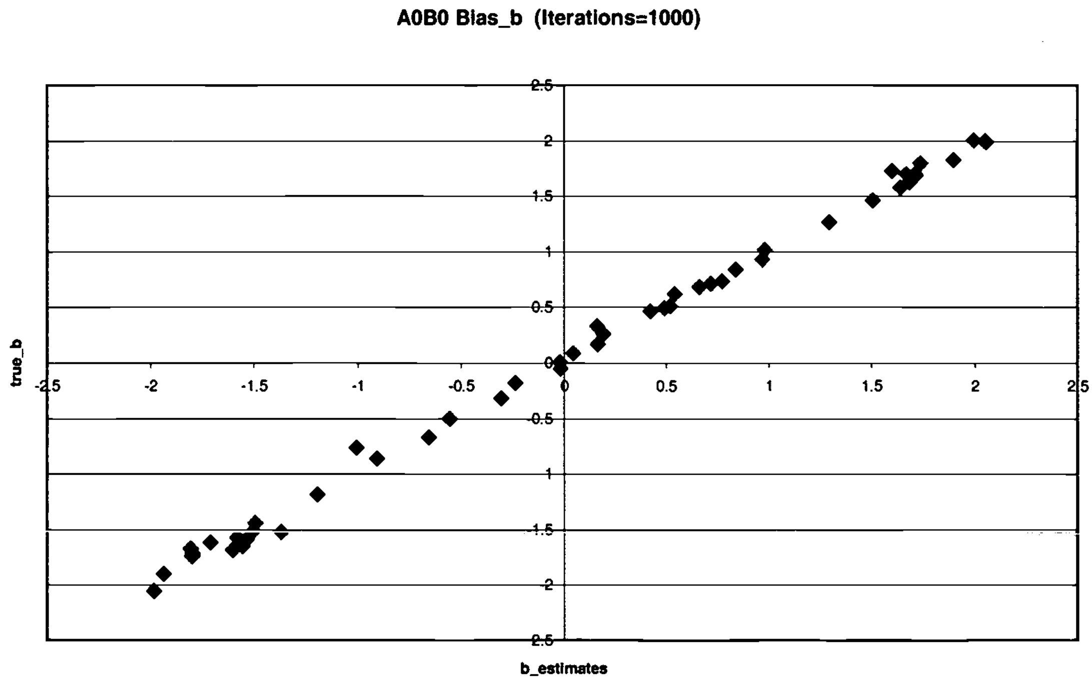
In an effort to gauge total bias (combined intentional and unintentional), and to shed light on part of the convergence issue, scatter plots based on 1,000 iterations were constructed. For purposes of this paper, bias is defined to be the difference between estimated and actual parameters.

Insert Tables 2 and 3 about here

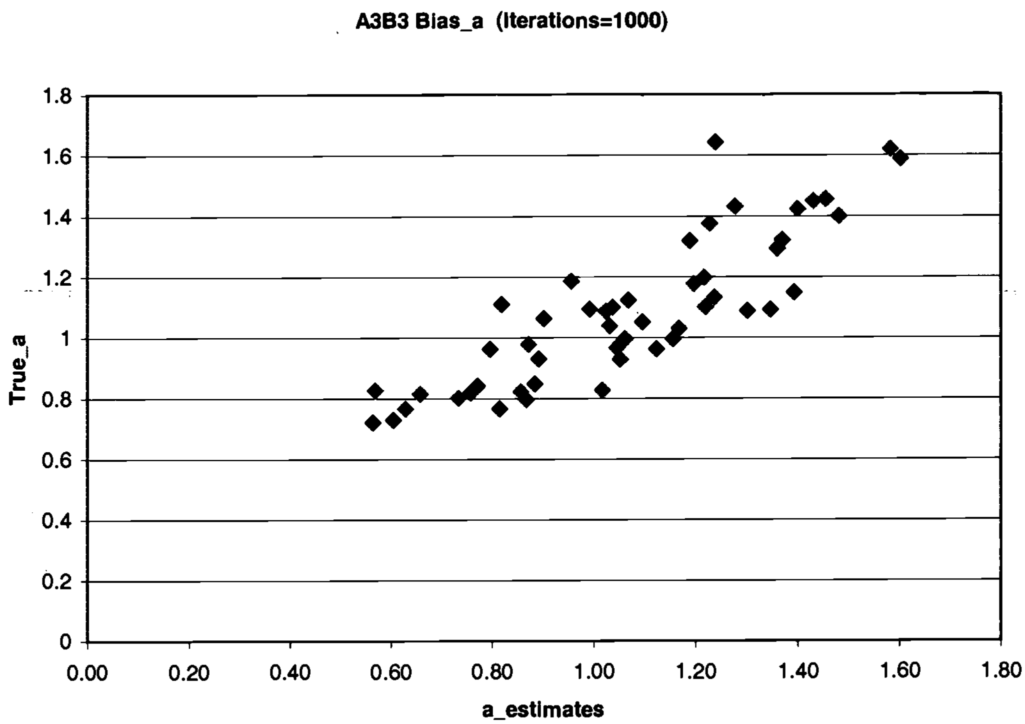
Scatter plots between estimated and actual values were produced for conditions A0B0 (our baseline condition) and A3B3. As the reader can see, baseline results are near-linear, as expected. **(Chart 6)**



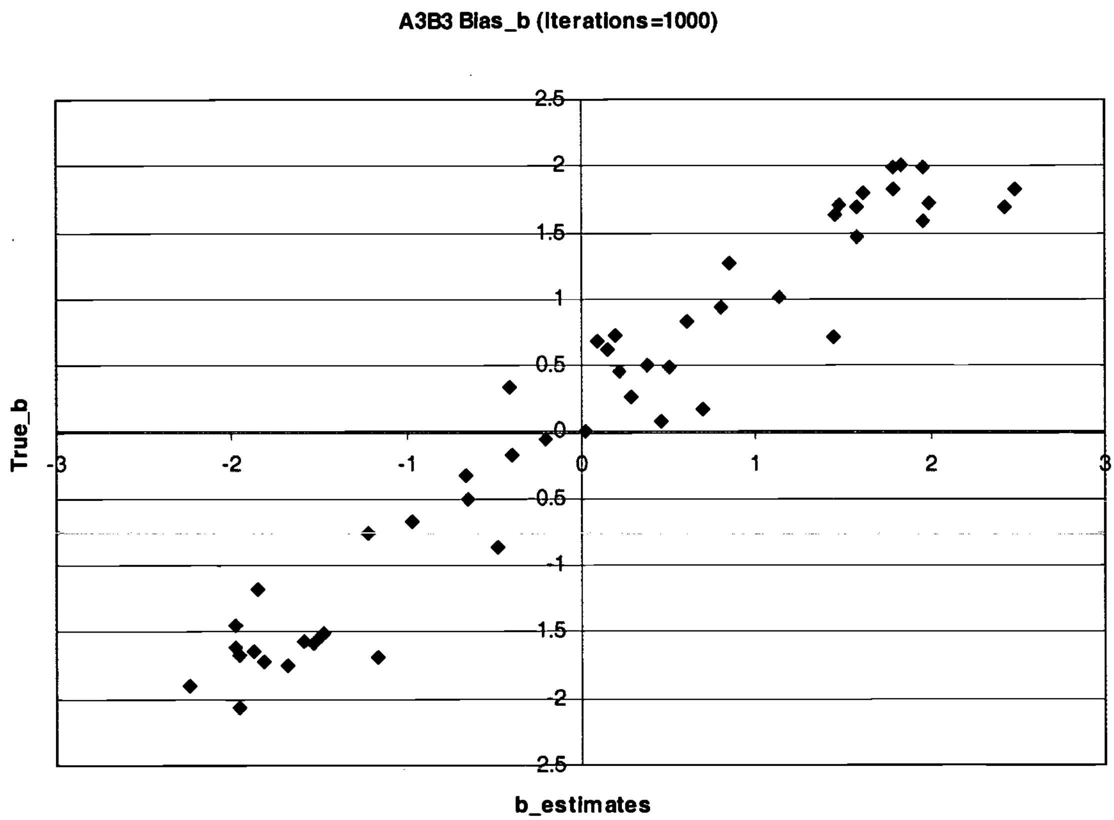
(Chart 7)


A0B0 Bias_b (Iterations=1000)
true_b
b_estimates
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5

Plots for condition a3b3 are located in charts 8 and 9. **(Chart 8)**



(Chart 9)



Upon inspection of these plots, it is possible to see a difference between the baseline condition A0B0 and the induced variance condition, A3B3, for both difficulty and discrimination. When comparing plots of A0B0-Bias_a with A3B3-Bias_a (and subsequently, A0B0-Bias_b with A3B3-Bias_b), an increase of scatter is evident.

A preliminary summary of bias findings is presented in Table 6.

Insert Chart 6 about here

Future Work

As mentioned above, data is currently being re-analyzed with significantly more cycles. It is impossible to foresee the optimal number of cycles needed for convergence. Since SCORIGHT is a new piece of analysis software, establishing the preferred number of cycles to achieve convergence is a trial-and-error process that has just begun. Once convergence is achieved with these data, it is expected that more accurate recovery of the generating values will be seen.

The ability of SCORIGHT, once established with simulation studies such as this one, to estimate the posterior distributions of item and ability parameters will provide new insight into item response data. Instead of relying solely on point estimates of parameters, the posterior sampling distributions will be estimated and could be used to modify standard estimation procedures in operational testing programs. In addition, SCORIGHT is designed to correctly estimate items that violate local independence assumptions that are believed to occur in assessment data. Items that rely on common stimuli such as reading passages, graphs, or tables of information and are dependent on each other have been shown to cause problems in standard estimation procedures (Worthington & Donoghue, 1997). Such items often are eliminated from operational assessment to achieve convergence of the item parameter estimates under typical estimation software. Use of SCORIGHT could resolve the loss of information that dropping dependent items from assessments inevitably causes.

References

Bejar, I. I. (1993). *A generative approach to psychological and educational measurement.* In N, Test theory for a new generation of tests (pp. 323-359). Hillsdale, NJ: Erlbaum Associates.

Kyllonen, P. C. (in press). Smart Testing. In R. Dillion (Ed.), Handbook on Testing. Westport, CT: Greenwood Publishing Group, Inc.

Meisner, R., Luecht, R., and Reckase, M. (1993). *The comparability of the statistical characteristics of test items generated by computer algorithms.* ACT Research Report Series 93-9. Iowa City, IA: American College Testing

Wang, X., Bradlow, E. T., and Wainer, H. (2000a). SCORIGHT [Computer software, v. 1.0]. Princeton, NJ: Educational Testing Service.

Wang, X., Bradlow, E. T., and Wainer, H. (2000b). *User's guide for SCORIGHT (version 1.2): A computer program for scoring tests built of testlets* (ETS RR-01-06). Princeton, NJ: Educational Testing Service.

Wang, X., Bradlow, E. T., and Wainer, H. (in press). *A general Bayesian model for testlets: theory and applications.* Applied Psychological Measurement.

Worthington, L., & Donoghue, J. R. (1997, March). *Detecting and describing local item dependence using IRT models: Lessons from the 1996 NAEP science assessment.* Paper presented at the Annual Meeting of the National Council on Measurement in Education. Chicago, IL.

Table 1

Generating Item Parameters

Item #	Discrimination	Difficulty	Guessing	Item #	Discrimination	Difficulty	Guessing
1	1.131	-2.058	.034	26	0.766	0.331	.261
2	1.050	-1.901	.078	27	1.448	0.460	.196
3	0.847	-1.745	.075	28	1.099	0.492	.278
4	1.197	-1.718	.199	29	1.100	0.509	.133
5	0.995	-1.685	.096	30	0.826	0.620	.246
6	0.994	-1.676	.236	31	1.292	0.679	.261
7	0.818	-1.649	.143	32	1.588	0.710	.111
8	0.766	-1.617	.209	33	0.826	0.733	.092
9	0.795	-1.586	.211	34	1.086	0.840	.182
10	1.086	-1.576	.178	35	0.801	0.932	.125
11	0.822	-1.527	.039	36	1.455	1.014	.301
12	0.962	-1.511	.135	37	1.320	1.266	.284
13	0.720	-1.444	.174	38	1.176	1.463	.108
14	1.185	-1.184	.202	39	1.029	1.577	.253
15	1.317	-0.862	.224	40	1.399	1.622	.259
16	0.814	-0.760	.316	41	0.977	1.682	.239
17	1.062	-0.672	.155	42	1.147	1.695	.152
18	0.9280	-0.502	.201	43	1.122	1.700	.134
19	1.422	-0.316	.094	44	1.037	1.726	.211
20	0.960	-0.179	.130	45	0.927	1.796	.232
21	1.092	-0.051	.165	46	1.644	1.821	.151
22	1.374	0.003	.091	47	0.841	1.822	.169
23	1.619	0.086	.152	48	0.965	1.984	.195
24	1.109	0.166	.177	49	0.729	1.996	.185
25	1.430	0.262	.175	50	1.091	2.004	.096

Table 2

Estimated and True Item Parameters: A0B0 (based on 1,000 iterations)

scaled_est_a	true_a	est_b	true_b	est_c	true_c	bias_a	bias_b	bias_c
1.193	1.131	-1.9825	-2.058	0.14	0.034	0.062	0.0755	0.106
1.073	1.05	-1.9369	-1.901	0.1683	0.078	0.023	-0.0359	0.0903
0.775	0.847	-1.7998	-1.745	0.1637	0.075	-0.072	-0.0548	0.0887
1.194	1.197	-1.7955	-1.718	0.1792	0.199	-0.003	-0.0775	-0.0198
1.072	0.995	-1.6017	-1.685	0.1574	0.096	0.077	0.0833	0.0614
1.020	0.994	-1.8063	-1.676	0.1483	0.236	0.026	-0.1303	-0.0877
0.866	0.818	-1.5565	-1.649	0.1965	0.143	0.048	0.0925	0.0535
0.743	0.766	-1.7074	-1.617	0.1828	0.209	-0.023	-0.0904	-0.0262
0.829	0.795	-1.5341	-1.586	0.2498	0.211	0.034	0.0519	0.0388
1.125	1.086	-1.5816	-1.576	0.166	0.178	0.039	-0.0056	-0.012
0.884	0.822	-1.3684	-1.527	0.162	0.039	0.062	0.1586	0.123
0.986	0.962	-1.5035	-1.511	0.1833	0.135	0.024	0.0075	0.0483
0.728	0.72	-1.4927	-1.444	0.151	0.174	0.008	-0.0487	-0.023
1.226	1.185	-1.1951	-1.184	0.2474	0.202	0.041	-0.0111	0.0454
1.322	1.317	-0.9092	-0.862	0.2217	0.224	0.005	-0.0472	-0.0023
0.752	0.814	-1.0067	-0.76	0.2113	0.316	-0.062	-0.2467	-0.1047
1.125	1.062	-0.6572	-0.672	0.1923	0.155	0.063	0.0148	0.0373
0.916	0.928	-0.5564	-0.502	0.19	0.201	-0.012	-0.0544	-0.011
1.381	1.422	-0.3048	-0.316	0.0996	0.094	-0.041	0.0112	0.0056
0.922	0.96	-0.2354	-0.179	0.1254	0.13	-0.038	-0.0564	-0.0046
1.079	1.092	-0.016	-0.051	0.187	0.165	-0.013	0.035	0.022
1.346	1.374	-0.0205	0.003	0.0936	0.091	-0.028	-0.0235	0.0026
1.640	1.619	0.0461	0.086	0.1458	0.152	0.021	-0.0399	-0.0062
1.120	1.109	0.1667	0.166	0.1574	0.177	0.011	0.0007	-0.0196
1.267	1.43	0.1947	0.262	0.1447	0.175	-0.163	-0.0673	-0.0303
0.759	0.766	0.1626	0.331	0.2456	0.261	-0.007	-0.1684	-0.0154
1.404	1.448	0.4207	0.46	0.1794	0.196	-0.044	-0.0393	-0.0166
1.064	1.099	0.4905	0.492	0.2974	0.278	-0.035	-0.0015	0.0194
1.143	1.1	0.5184	0.509	0.1615	0.133	0.043	0.0094	0.0285
0.849	0.826	0.5415	0.62	0.2414	0.246	0.023	-0.0785	-0.0046
1.307	1.292	0.6614	0.679	0.2567	0.261	0.015	-0.0176	-0.0043
1.579	1.588	0.7175	0.71	0.1195	0.111	-0.009	0.0075	0.0085
0.857	0.826	0.7756	0.733	0.1162	0.092	0.031	0.0426	0.0242
1.030	1.086	0.8405	0.84	0.1812	0.182	-0.056	0.0005	-0.0008
0.711	0.801	0.9661	0.932	0.1373	0.125	-0.090	0.0341	0.0123
1.397	1.455	0.979	1.014	0.3008	0.301	-0.058	-0.035	-0.0002
1.211	1.32	1.2958	1.266	0.2849	0.284	-0.109	0.0298	0.0009
1.160	1.176	1.5059	1.463	0.128	0.108	-0.016	0.0429	0.02
1.158	1.029	1.6409	1.577	0.2699	0.253	0.129	0.0639	0.0169
1.267	1.399	1.6853	1.622	0.2568	0.259	-0.132	0.0633	-0.0022
0.963	0.977	1.7133	1.682	0.2424	0.239	-0.014	0.0313	0.0034
1.010	1.147	1.6692	1.695	0.1342	0.152	-0.137	-0.0258	-0.0178
1.196	1.122	1.7132	1.7	0.1339	0.134	0.074	0.0132	-0.0001
1.126	1.037	1.6	1.726	0.2117	0.211	0.089	-0.126	0.0007
0.859	0.927	1.7373	1.796	0.2217	0.232	-0.068	-0.0587	-0.0103
1.312	1.644	1.8941	1.821	0.1514	0.151	-0.332	0.0731	0.0004

0.802	0.841	1.8988	1.822	0.1642	0.169	-0.039	0.0768	-0.0048
0.914	0.965	2.0551	1.984	0.1992	0.195	-0.051	0.0711	0.0042
0.625	0.729	2.0501	1.996	0.169	0.185	-0.104	0.0541	-0.016
1.227	1.091	1.9957	2.004	0.1053	0.096	0.136	-0.0083	0.0093

Table 3

Estimated and True Item Parameters: A3B3 (based on 1,000 iterations)

scaled_est_a	true_a	est_b	true_b	est_c	true_c	bias_a	bias_b	bias_c
1.2368	1.131	-1.954	-2.058	0.1818	0.034	0.1058	0.104	0.1478
1.0955	1.05	-2.245	-1.901	0.2174	0.078	0.0455	-0.3443	0.1394
0.8845	0.847	-1.683	-1.745	0.1283	0.075	0.0375	0.0624	0.0533
1.2165	1.197	-1.824	-1.718	0.1791	0.199	0.0195	-0.106	-0.0199
1.0612	0.995	-1.173	-1.685	0.1818	0.096	0.0662	0.5122	0.0858
1.1549	0.994	-1.955	-1.676	0.1793	0.236	0.1609	-0.2789	-0.0567
0.7568	0.818	-1.88	-1.649	0.1556	0.143	-0.0612	-0.2307	0.0126
0.8139	0.766	-1.976	-1.617	0.177	0.209	0.0479	-0.3593	-0.032
0.8673	0.795	-1.537	-1.586	0.2405	0.211	0.0723	0.0494	0.0295
1.3005	1.086	-1.593	-1.576	0.1599	0.178	0.2145	-0.0168	-0.0181
0.8565	0.822	-1.484	-1.527	0.1993	0.039	0.0345	0.0427	0.1603
0.7954	0.962	-1.473	-1.511	0.1404	0.135	-0.1666	0.0384	0.0054
0.5643	0.72	-1.984	-1.444	0.1674	0.174	-0.1557	-0.5395	-0.0066
0.9554	1.185	-1.854	-1.184	0.1556	0.202	-0.2296	-0.6696	-0.0464
1.1882	1.317	-0.476	-0.862	0.1678	0.224	-0.1288	0.386	-0.0562
0.6571	0.814	-1.223	-0.76	0.233	0.316	-0.1569	-0.4628	-0.083
0.9015	1.062	-0.976	-0.672	0.1144	0.155	-0.1605	-0.3036	-0.0406
0.8921	0.928	-0.656	-0.502	0.1881	0.201	-0.0359	-0.1544	-0.0129
1.3986	1.422	-0.666	-0.316	0.0867	0.094	-0.0234	-0.3496	-0.0073
1.1215	0.96	-0.404	-0.179	0.159	0.13	0.1615	-0.2254	0.029
0.9911	1.092	-0.21	-0.051	0.1716	0.165	-0.1009	-0.1587	0.0066
1.2270	1.374	0.0241	0.003	0.0716	0.091	-0.1470	0.0211	-0.0194
1.5816	1.619	0.4612	0.086	0.1644	0.152	-0.0374	0.3752	0.0124
0.8182	1.109	0.6973	0.166	0.1592	0.177	-0.2908	0.5313	-0.0178
1.2762	1.43	0.2852	0.262	0.1532	0.175	-0.1538	0.0232	-0.0218
0.6283	0.766	-0.408	0.331	0.2876	0.261	-0.1377	-0.7385	0.0266
1.4299	1.448	0.2166	0.46	0.2009	0.196	-0.0181	-0.2434	0.0049
1.0366	1.099	0.5086	0.492	0.2698	0.278	-0.0624	0.0166	-0.0082
1.2189	1.1	0.38	0.509	0.1465	0.133	0.1189	-0.129	0.0135
0.5686	0.826	0.1474	0.62	0.1643	0.246	-0.2574	-0.4726	-0.0817
1.3589	1.292	0.0879	0.679	0.2578	0.261	0.0669	-0.5911	-0.0032
1.6019	1.588	1.4448	0.71	0.1049	0.111	0.0139	0.7348	-0.0061
1.0169	0.826	0.1962	0.733	0.0909	0.092	0.1909	-0.5368	-0.0011
1.0235	1.086	0.6099	0.84	0.1563	0.182	-0.0625	-0.2301	-0.0257
0.7327	0.801	0.8061	0.932	0.1117	0.125	-0.0683	-0.1259	-0.0133
1.4545	1.455	1.1307	1.014	0.302	0.301	-0.0005	0.1167	0.001
1.3691	1.32	0.8428	1.266	0.2696	0.284	0.0491	-0.4232	-0.0144
1.1958	1.176	1.5775	1.463	0.1163	0.108	0.0198	0.1145	0.0083

1.1666	1.029	1.9627	1.577	0.2554	0.253	0.1376	0.3857	0.0024
1.4799	1.399	1.4597	1.622	0.2598	0.259	0.0809	-0.1623	0.0008
0.8715	0.977	2.4275	1.682	0.2376	0.239	-0.1055	0.7455	-0.0014
1.3919	1.147	1.5818	1.695	0.1423	0.152	0.2449	-0.1132	-0.0097
1.0671	1.122	1.4772	1.7	0.1255	0.134	-0.0549	-0.2228	-0.0085
1.0311	1.037	1.9912	1.726	0.2083	0.211	-0.0059	0.2652	-0.0027
1.0517	0.927	1.6159	1.796	0.2321	0.232	0.1247	-0.1801	1E-04
1.2392	1.644	1.7835	1.821	0.1433	0.151	-0.4048	-0.0375	-0.0077
0.7706	0.841	2.4884	1.822	0.1517	0.169	-0.0704	0.6664	-0.0173
1.0449	0.965	1.7872	1.984	0.2	0.195	0.0799	-0.1968	0.005
0.6046	0.729	1.9537	1.996	0.2196	0.185	-0.1244	-0.0423	0.0346
1.3456	1.091	1.8285	2.004	0.1075	0.096	0.2546	-0.1755	0.0115

Table 6

This table summarizes the number of items (out of 50) that exhibited a notable level of bias.

Let Bias = Expected – Actual
Flagging criteria for bias:

$\Delta a \exists 0.25$, where $\Delta a = a$_estimate – true_a

$\Delta b \exists 0.50$, where $\Delta b = b$_estimate – true_ b

	A0B0	**A3B3**
1,000 Iterations	a: 1/50, b: 0/50	a: 4/50, b: 10/50
25,000 Iterations	Result pending	Result pending

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM032877

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Modeling the Hyperdistribution of Item Parameters to Improve the Accuracy of Recovery in Estimation Procedures

Author(s): Joy L. Matthews-López / Catherine M. Hombo

Corporate Source: Educational Testing Service

Publication Date: April, 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Joy L. Matthews-López

Printed Name/Position/Title: Joy L. Matthews-López, measurement statistician

Organization/Address: Educational Testing Service, Rosedale Road, MS 31-L, Princeton, NJ 08541

Telephone: (609) 683-2208

FAX: (609) 683-2130

E-Mail Address: jmatthews-lopez@ets.org

Date: 4/22/01

or (740) 594-4600

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com