DOCUMENT RESUME

ED 269 476 TM 860 334

AUTHOR	Buhr, Dianne C.; Algina, James
TITLE	A Comparison of Item Parameter Estimates and Ability Parameter Estimates Obtained by Different Methods Implemented by BILOG.
SPONS AGENCY	Florida Univ., Gainesville. Inst. for Student Assessment and Evaluation.
PUB DATE	. >r 86
NOTE	38p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Ability; *Bayesian Statistics; Comparative Analysis; Computer Oriented P ograms; Correlation; *Estimation (Mathematics); *Item Analysis; *Latent Trait Theory; Mathematical Models; Maximum Likelihood Statistics; Statistical Distributions; Statistical Studies
IDENTIFIERS	Ability Estimates; Ability Parameters; *BILOG Computer Program; *Item Parameters; One Parameter Model; Robustness; Three Parameter Model; Two-Parameter Model

ABSTRACT

The focus of this study is on the estimation procedures implemented in BILOG, a computer program. One purpose is to compare the item parameter estimates produced by various procedures available in BILOG. Four different models are used: the one, two, and three parameter model and a three parameter model with common guessing parameters. The results generally indicate that the various item parameter estimation procedures tend to yield similar results. The major exception concerned the Bayesian and maximum likelihood procedures (MLP) applied to the three parameter model. The MLP has a tendency to produce more extreme estimates than the Bayesian procedure. A second purpose is to compare the ability estimates produced by the available procedures: maximum likelihood, expected a'posteriori, and maximum a'posteriori. The results indicated that: (1) robustification is not a strong effect on the mean or standard deviation of the ability estimates; (2) the mean and variance of the ability estimates are not strongly effected by the type of item parameter estimates used in calculating ability estimates; and (3) the effect of ability estimation procedure is fairly strong on the ability estimates. (PN)

A COMPARISON OF ITEM PARAMETER ESTIMATES AND ABILITY PARAMETER ESTIMATES OBTAINED BY DIFFERENT METHODS IMPLEMENTED BY BILOG

Dianne C. Buhr
James Algina

University of Florida

A paper presented at the annual meeting
of AREA, San Francisco, 1986

A Comparison of Item Parameter Estimates and of Ability Parameter Estimates Obtained By Different Methods Implemented by BILOG[1]

A binary item response theory (IRT) model is a model for the relationship between binary scores on a test item and scores on a latent (or unobservable) trait. The curve expressing the relationship is called an item characteristic curve (ICC). The most popular binary IRT models are the normal ogive model and the one, two, and three parameter logistic models. The development of procedures for estimating the parameters of binary IRT models has a history dating back about 50 years. As Baker (1977) notes, initial attempts to solve the estimation problem generally involved substituting an observed score, usually a total score on the test, for the latent trait and estimating the item parameters of each ICC independently. Finney (1944) presented this kind of maximum likelihood estimation procedure for estimating parameters of the normal ogive model. Earlier Richardson (1936), Ferguson (1942), and Lawley (1943) had applied the constant process, a generalized least squares procedure, to the estimation task. The maximum likelihood and constant process approach typically yield similar estimates (Baker, 1965) though the former can encounter problems when a score group has a large proportion of examinees answering correctly or incorrectly (Finney, 1944). Both approaches can be applied to the logistic models.

Another approach to estimating item parameters of the normal ogive is based on Richardson's (1936) demonstration of the functional relationship between the item parameters of the normal ogive model, and the item-latent trait correlation and item difficulty of classical test theory. When the item-observed score correlation is substituted for the item-latent trait correlation, an approximate procedure is obtained for estimating the normal ogive parameters. More recently, Urry (1974) extended this procedure to the three parameter normal ogive model.

In recent years, three new maximum likelihood procedures have become available. These procedures, which do not require substituting an observed score for the latent trait score, are the conditional (CML), joint (JML), and marginal maximum likelihood (MML) procedures. All three can be applied to the one parameter logistic (Rasch) model. The latter two can be applied to the two and three parameter logistic models. There are computer programs available to implement each of the CML, JML, and MML procedures with the logistic models. The PML (Gustafsson, 1977) program implements the CML procedure for the one parameter logistic model. BICAL (Wright, Mead and Bell, 1979) implements the JML procedure for the Rasch model. LOGIST 5 (Wingersky, Barton, and Lord, 1982) calculates JML estimates for all three logistic models, whereas BILOG (Mislevy & Bock, 1982) implements the MML for the three logistic models. Swaminathan (1983) gave a detailed presentation of the three types of estimators.

CML and MML item-parameters estimators are consistent estimators. This may be a significant advantage over the JML estimators which are inconsistent when the number of items is finite. However, for the one parameter logistic model the JML estimators have been shown to be consistent as the number of items and examinees each tend to infinity (Haberman, 1975). Empirical results (Lord, 1975; Swaminathan and Gifford, 1983) suggest this result may hold for the two and three parameter logistic models.

There has also been some interest in Bayesian estimation of the item parameters of the logistic models. Swaminathan and Gifford (1982) developed a Bayesian procedure for use with the one parameter logistic model. The Bayesian procedure has been extended to the two parameter logistic model by Swaminathan and Gifford (1985). BILOG implements a Bayesian procedure for all three models. However, it differs in several details from the Swaminathan-Gifford procedure.

Just as there are several procedures available for estimating the item parameters, there are several for estimating the ability parameters. The name of the JML procedure derives from the fact that it simultaneously estimates the item and ability parameters. Thus there are JML estimators of ability parameters. Similarly the Swaminathan-Gifford Bayesian procedures simultaneously estimate both the item and ability parameters. Other ability estimation procedures assume the item parameters are known. Both a maximum likelihood procedure and a variety of Bayesian procedures are available. BILOG incorporates both kinds of procedures. However, of course, in practice these procedures are implemented using estimates of the item parameters.

Purpose Of The Study

The focus of this study is on the estimation procedures implemented in BILOG. One purpose is to compare the item parameter estimates produced by various procedures available in BILOG. Four different models were used: the one, two, and three parameter model and a three parameter model with common guessing parameters. For item parameter estimation, BILOG basically implements the MML and Bayesian procedures. However, th options available in the program give the user a fairly wide set of choices about the implementation of the procedures. These will be described in the succeeding section.

A second purpose is to compare the ability estimates produced by the available procedures: maximum likelihood, expected a' posteriori, and maximum a' posteriori. The latter two are Bayesian procedures. For each of the three procedures biweight robustification is the only available option. The effect of robustification was investigated.

Marginal Maximum Likelihood Procedure

In this section we describe the MML procedure in the context of the three parameter logistic model. Let $P_i(\Theta_j)$ denote the probability that the jth examinee (j=1,...,n) answers the ith item correctly (i=1,...,N). Let u_{ij} be a binary variable. For a correct response $u_{ij}=1$; for an incorrect response $u_{ij}=0$. Let $\underline{\Theta}' = [\Theta_1 \ldots \Theta_n]$ be the vector of latent trait scores, and let $\underline{a}' = [a_1 \ldots a_N]$, $\underline{b}' = [b_1 \ldots b_N]$, and $\underline{c}' = [c_1 \ldots c_N]$ be vectors of item discrimination, difficulty, and guessing parameters respectively.

For the jth examinee the likelihood function for the data is

$$L(\underline{u}_j|\Theta_j\underline{a},\underline{b},\underline{c})= \prod_i [P_i(\Theta_j)]^{u_{ij}}[1-P_i(\Theta_j)]^{1-u_{ij}}$$

where $\underline{u}'_j=[u_{1j}\ldots u_{Nj}]$. The notation emphasizes that the likelihood function is conditioned on the jth ability parameter and the item parameters for all items. For all N examinees the likelihood function is

$$L(\underline{u}|\Theta_j,\underline{a},\underline{b},\underline{c})= \prod_j L(\underline{u}_j|\Theta_j,\underline{a},\underline{b},\underline{c})$$

where $\underline{u}'=[\underline{u}'_1\ldots\underline{u}'_n]$. The JML procedure simultaneously computes the $\underline{\Theta}$, $\underline{a}$, $\underline{b}$, and $\underline{c}$ that maximize the latter likelihood function. Thus n+3N parameters are estimated for the three parameter model.

In the MML procedure each examinee's latent trait score (Θ) is considered to be randomly chosen from a population with ability distribution $f(\Theta)$. The marginal likelihood of the data for the jth examinee is

$$L(\underline{u}_j|\underline{a},\underline{b},\underline{c})= \int L(\underline{u}_j|\Theta_j,\underline{a},\underline{b},\underline{c})f(\Theta)d\Theta$$

Essentially, the marginal likelihood is obtained as a weighted average of the conditional likelihoods

$$L(\underline{u}_j|\Theta_j,\underline{a},\underline{b},\underline{c})$$

where the weights are determined by $f(\Theta)$. This weighting process removes the dependence on Θ and, therefore, in the process of estimating item parameters, it is not necessary to estimate the ability parameters for the N examinees. The function maximized in the MML procedure is

$$L(\underline{u}|\underline{a},\underline{b},\underline{c}) = \prod_j L(\underline{u}_j|\underline{a},\underline{b},\underline{c})$$

To implement the MML procedure it is necessary to make an assumption about the form of $f(\Theta)$. In BILOG the default option is for $f(\Theta)$ to be a standard normal distribution. However, the program permits the user to specify other distributions.

In BILOG, the distribution $f(\Theta)$ can be treated in either of two ways. In one, $f(\Theta)$ is treated as a distribution to be estimated. Thus the assumed $f(\Theta)$ is the basis for starting values in an iterative procedure for estimating $f(\Theta)$ and the item parameters. As Mislevy and Bock (1982) note this type of procedure is similar to the JML procedure. We refer to it as marginal maximum likelihood with estimation of ability distribution MML-EAD. In the other, $f(\Theta)$ is treated as an assumption about the distribution of latent ability. The same distribution is employed throughout the iterative procedure for estimating the item parameters. This is the MML procedure. Both options were investigated in the study. We investigated three forms for $f(\Theta)$. Two were a normal distribution and a uniform distribution, each with mean zero and standard deviation one. For the third, we used BILOG to estimate $f(\Theta)$ on one sample and then used this estimate of $f(\Theta)$ in applying the MML procedure to a second sample. Both samples were chosen randomly from a larger sample.

Bayesian Procedures

The Bayesian procedures incorporate assumptions about the distributions of item parameters. These assumed distributions are called prior distributions. The default prior distributions employed in BICAL are: normal, with mean zero and standard deviation two, for the difficulty parameters; lognormal, with mean $e^{.5}$ and variance $e^{1}(e^{1}-1)$ for the discrimination parameters, and beta, with $\alpha=20p+1$ and $\beta=20(1-p)+1$. For the beta distribution, p is the reciprocal of the number of alternat. ves.

The incorporation of the prior distributions into the estimation procedure makes it unlikely for the estimates to occur in regions that are less probable according to the prior distribution. For example, the default prior distribution for difficulty parameters is a normal distribution with mean zero and standard deviation 2. Thus difficulty estimates <-2 or >2 are substantially less likely than estimates between - 2 and 2. Difficulty estimates <-4 or >4 are very unlikely to occur. The default prior distributions in BILOG are relatively diffuse. That is, they do not constrain the estimates to unreasonably small regions of the parameter space. However BILOG permits the user to tailor the priors to the specific application. Thus the user can use more diffuse priors or tighter priors. In addition the user can also choose which parameters to place priors on. In the present study we implemented the Bayesian procedures by employing default priors on all item parameters.

When using the Bayesian procedure in BILOG the user can either specify that the priors remain the same at each iteration or that the parameters of the priors be updated on each interation. We refer to the former as Bayesian estimation (BE) and the latter as Bayesian estimation with updating of item priors (BE-UIP). For the kth iteration, the updating of the prior distribution of the item difficulties, for example, consists of substituting the mean of the item difficulty estimates from the (k-1)th iteration for the mean of the assumed prior distribution (which equals zero in the default prior for item difficulties). The updating of the other priors also involves substitution of the appropriate mean parameter estimate from the (k-1)th iteration. When we employed prior distributions, we investigated the effect of updating on the parameter estimates.

Other Options for Item Parameter Estimation

In addition to the options described above, with the three parameter model and the three parameter model with common guessing parameters there are three options for treatment of omitted items. Omitted items can be treated as incorrect, not presented, or fractionally correct. Mislevy and Bock (1984) pointed out that the second option permits an examinee to obtain high scores by responding only to items the examinee is sure of.

This optior ould seem to favor the extremely cautious examinee and it was not investigated.

Method

Instrument

The test used in the study had 39 items and was relatively easy. The mean and standard deviation, for number correct scores, were 31.3 and 7.8. Frequency distributions for number correct scores, proportion-correct item difficulties, and item-total point biserials are displayed in Table 1.

Design

The levels of the factors in the design for investigating parameter estimation procedures were:

1. Model-one parameter, two parameter, three parameter, and three parameter with common guessing parameters;
2. Sample size - 250, 500, 750, and 1000 examinees;
3. Ability distributions - normal, uniform, and empirical;
4. Estimation procecures: MML, MML-EAD, BE, BE-UIP;
5. Scoring of omits: incorrect and fractionally correct for the three parameter and three parameter common c models. For the one and two parameter model, only incorrect scoring of omits is implemented in BILOG.

Not all possible condition combinations were investigated. In particular the effect of ability distribution was not investigated with samples of 500 or 750. With samples of 1000, the effect of ability distribution was only investigated in connection with the three parameter model and the three parameter common c model.

Three methods of estimating ability parameters were investigated: maximum likelihood (ML), maximum a' posteriori (MAP), and expected a' posteriori (EAP). In addition the effect of biweight robustification was investigated. To implement each ability estimation procedure, a set of item parameter estimates is required. MML and BE estimates, obtained using both normal and uniform ability distributions, were employed. This choice of item parameter estimates was based on the results of the comparison of item parameter estimates. Only the three parameter model was employed, and only item parameter estimates based on samples of 250 were employed. Again these decisions were based on the comparisons of the item parameter estimates.

- - - - - - - - - - - - - -

Insert Table 1 About Here

- - - - - - - - - - - - - -

Results

Ability Distribution

A normal, an empirical, and a uniform ability distribution were employed in the MML procedure to obtain three sets of parameter estimates for the three parameter model. For a particular sample size, the same sample was used with each prior distribution. Means and standard deviations for each set of estimates, based on a sample of 250, are reported in Table 2. Also reported are correlations between the a_g's , between the

b_g's, and between the c_g's estimated using the three ability distributions. The results indicate that the estimates based on the normal and empirical distributions are quite similar. There is less similarity between the estimates based on the normal and uniform distributions and between the estimates based on the empirical and uniform distributions. Nevertheless the agreement is still quite substantial.

- - - - - - - - - - - - -

Insert Table 2 About Here

- - - - - - - - - - - - -

The effect of the ability distribution on item-parameter estimates for the three parameter model was also examined in connection with three other estimation schemes: MML-EAD, BE, and BE-UIP. The results for the MML-EAD procedure were very similar to those reported in Table 2. The results for the BE and BE-UIP were also quite similar to one another. Results based on the BE procedure are reported in Table 3. Comparing the results in Table 3 to those in Table 2 indicates the Bayesian procedures were even less affected than the marginal maximum-likelihood procedures were by the choice of the ability distribution.

- - - - - - - - - - - - -

Insert Table 3 About Here

- - - - - - - - - - - - -

The effect of ability distribution on item-parameter estimates was also investigated in connection with the one and two parameter model, and the three parameter common c models. With these simpler models, the effect of the ability distribution was as small or smaller than it was for the three parameter model. This trend is illustrated by the results, reported in Table 4, for the MML procedure applied with the three-parameter common c model.

- - - - - - - - - - - - -

Insert Table 4 About Here

- - - - - - - - - - - - -

The preceding results are based on estimates obtained by scoring omits as wrong answers. With the three-parameter model (with or without a common c) omits can also be scored as fractionally correct. With this latter option, the effects of ability distributions were small and approximately the same as with the former option.

For all of the preceding results, the sample size was 250. It seemed unlikely that the effect of the prior ability distribution would increase as the sample size increased. However, to check this possibility, a sample size of 1000 and a normal, an empirical, and a uniform distribution were employed with each of the four estimation procedures applied to the three parameter model and the three parameter common c model. Omits were scored as incorrect. For the MML procedure, means, standard deviations, and correlations are reported in Table 5. Comparison of the results in Tables 2 and 5 indicates that the effect of ability distribution is independent of the sample size. The effect of sample size was similar for the other estimation procedures and for the three parameter common c model.

- - - - - - - - - - - - -

Insert Table 5 About Here

- - - - - - - - - - - - -

Type of Estimation Procedure

Four different estimation procedures (MML, MML-EAD, BE, and BE-UIP) were employed to obtain four sets of estimates for the three parameter model. Means and standard deviations for each set of estimates, based on a sample of 250 and a normal prior, are reported in Table 6. Also reported are correlations between the a_g's, between the b_g's, and between the c_g's obtained by using the four methods. The results indicate that the two MML procedures yield similar estimates as do the two Bayesian procedures. However between the two types of procedures (marginal maximum likelihood and Bayesian), the estimates are less similar.

- - - - - - - - - - - - -

Insert Table 6 About Here

- - - - - - - - - - - - -

The effect of estimation procedure was also investigated with three simpler models: the one and two parameter models, and the three parameter common c model. Estimation procedure had almost no effect with the simpler models. This is illustrated by results for the three parameter common c model reported in Table 7. The estimates described by these results were calculated using a normal ability distribution. With the empirical and uniform ability distributions the results for the three parameter common c model were also unaffected by method of estimation.

Similarly, with each ability distribution the results for the one and two parameter models also indicated a lack of effect for estimation procedure

- - - - - - - - - - - - - -

Insert Table 7 About Here

- - - - - - - - - - - - - -

The preceding results are based on a sample size of 250. The effect of method of estimation was also investigated with a normal ability distribution and samples of 500, 750, and 1000 examinees. For a sample size of 250 and with the three simple models, estimation-method effect was quite small. With larger sample sizes it appeared to become even smaller. With the three parameter model the effect of sample size depended on the parameter. For the a_g parameter, increasing the sample size from 250 to 500 increased the between estimation-method correlations and decreased the between method differences in means and standard deviations. Further increases in sample size appeared not to effect the similarity of the estimates. These trends are shown in Table 8. For the b_g parameter, increasing the sample size had a negligible effect on the between estimation-method correlations. The between method differences in means tended to decrease as the sample size increases from 250 to 500 and remain about the same with further increases. The effect of increased sample size on between method differences in standard deviations was irregular. The between method differences in standard deviations increased as the sample size changed to 500 then decreased as the sample size increased to 750 and decreased again as the sample size increased to 1000 examinees. This trend principally reflects the behavior of the largest of the estimated b_g's. With the MML procedure, for example, the largest of the estimated b_g's were approximately -4, -12, -8, and -5 with 250, 500, 750, and 1000 examinees respectively. Thus the trend found for standard deviations may not occur with other tests. For the c_g parameter, there were negligible effects of increasing sample size on between method differences in parameter estimates.

- - - - - - - - - - - - - -

Insert Table 8 About Here

- - - - - - - - - - - - - -

As expected, the maximum likelihood procedures tended to result in more extreme estimates than the Bayesian procedures. This tendency was most marked with three parameter model and is illustrated in Table 9. The results in Table 9 describe the estimates obtained using a normal ability distribution. The tendency for the maximum likelihood procedures to produce extreme estimates was not reduced by using the empirical or the uniform ability distribution. With the simpler models the maximum likelihood procedures had less of a tendency to produce extreme estimates. When extreme estimates were produced the discrepancies between the maximum likelihood and Bayesian estimates tended to be smaller than they were with the three parameter model.

- - - - - - - - - - - - - -

Insert Table 9 about Here

- - - - - - - - - - - - - -

Scoring of Omits

For the three parameter model and the three parameter common c model, parameters were estimated with omits scored as wrong and omits scored as fractionally correct. The effect of the method of scoring was relatively minor for the sample of 250 and decreased with increasing sample size. This trend is illustrated in Table 10 which contains results for the three parameter model and MML parameter estimation.

- - - - - - - - - - - - - - - -

Insert Table 10 About Here

- - - - - - - - - - - - - - - -

Ability Estimates

Correlations among the ML, EAP, and MAP ability estimates, based on MML and BE item parameter estimates obtained using a sample of 250 examinees and a normal ability distribution, are reported in Table 11. The sample size for the correlations was also 250 and the sample was the same as the one used to obtain the item parameter estimates. The correlations are all above .90. Similar results were obtained for ability estimates based on MML and BE item parameter estimates obtained using a uniform ability distribution. The cross correlations between the two sets of ability estimates were also all above .90.

Insert Table 11 About Here

Means and standard deviations for the ability estimates, calculated using item parameters estimates based on a normal ability distribution, are reported in Table 12. The item estimation procedures - MML and BE - had a relatively small effect on the mean ability estimate. Controlling for ability estimation procedure and robustification, the mean differences range in absolute value from .03 and .06. Similarly, the effect on standard deviations was small. The effect of robustification was relatively small; it increased the mean ability estimate, with increases between .03 to .06. The effect on standard deviations was also small. The effect of ability estimation procedure - ML, EAP, and MAP - on means and standard deviation was relatively large. Moreover the effect appears to be larger with robustification than without. In general, the ML estimates had the largest mean and standard deviation. The MAP estimates had the smallest mean and standard deviation. The effect was due to the fact that the ML procedure produced much higher maximum ability estimates than either the EAP or MAP. Minimum and maximum ability estimates are shown in Table 13. Both the differences in the mean estimates and in the maximum estimates are of sufficient size to be of practical significance. This is particularly true for the differences between the ML estimates and either the MAP or the EAP estimates.

- - - - - - - - - - - - - - - - -

Insert Tables 12 and 13 About Here

- - - - - - - - - - - - - - - - -

As noted earlier, the ability estimates were obtained using item parameter estimates that were calculated using a sample size of 250. Because the item parameters are treated as known in the ability estimation phase, increasing the sample size in the item parameter estimation phase should not have any impact on the effect ability estimation procedure has on the ability estimates.

The general trend in the results were the same for ability estimates based on item parameter estimates calculated using the uniform distribution. The effect of robustification was of about the same magnitude as in the preceding results. The effect of item estimation procedure - MML or BE - on mean ability estimates was, however, larger. It ranged in absolute value from .09 to .13. The effect of ability estimation procedure - ML, EAP, or MAP - was about the same magnitude as in the preceding results.

Summary

The results indicate that, for the most part, the various item parameter estimation procedures tend to yield similar results. The major exception to this generalization concerned the Bayesian and maximum likelihood procedures applied to the three parameter model. With 250 examinees, correlations between a_g estimates averaged about .75 for maximum likelihood-Bayesian pairs of estimation procedures. For c_g estimates the correlation was likewise about .75. For the b_g estimates the correlations averaged about .92. For the a_g estimates, these correlations increased to between .90 and .95 with sample sizes of 500, 750, and 1000 examinees. The correlations for the b_g and c_g estimates were largely unaffected by changes in the sample sizes.

The maximum likelihood procedure had a tendency to produce more extreme estimates than the Bayesian procedure. This tendency was most pronounced when the procedures were applied to the three parameter model.

Ability estimation was investigated only in connection with the three parameter model. Generally, the correlations were high between ability parameter estimates obtained using the various approaches studied in this research. In addition, the results indicated that robustification did not strongly effect the mean or standard deviation of the ability estimates. The results also indicated that the mean and variance of the ability estimates were not strongly effected by the type of item parameter estimates used in calculating the ability estimates, at least when the item parameter estimates were based on a normal ability distribution. The effect of type of item parameter estimate was stronger when the item parameters were calculated using a uniform ability distribution. The importance of the latter finding is that it emphasizes the possibility that the former results may be sample and/or test specific. In studying the effect of item estimation procedure, the sample for which ability estimates were obtained was also used for item parameter estimation. It is possible that the method of item parameter estimation might have a stronger effect when ability estimates are calculated for a new sample. This should be investigated.

There was a fairly strong effect of ability estimation procedure on the ability estimates. The largest discrepancies were between the ML procedure, on one hand, and the EAP and MAP procedures on the other. Additional research should be undertaken to determine whether these differences will also occur with the simpler models. In addition analyses of simulated data should be undertaken to determine whether any of the three procedures produces substantially biased ability estimators.

Footnote

This research was supported by a grant from the Institute for Student Assessment and Evaluation, University of Florida.

References

Baker, F. B. (1965). Origins of X_{50} and β as a modern item analysis technique. Journal of Educational Measurement, 2, 167-180.

Baker, F. B. (1977). Advances in item analysis. Review of Educational Research, 47, 151-178.

Ferguson, G. A. (1942). Item selection by the constant process. Psychometrika, 7, 19-29.

Finney, D. J. (1944). The application of probit analysis to the results of mental tests. Psychometrika, 9, 31-39.

Gustaffson, J. E. (1977). The Rasch model for dichotomous items: Theory, applications and a computer program. (Reports from the Institute of Education, No. 63). Goteborg: University of Goteborg.

Haberman, S. (1975). Maximum likelihood estimates in exponential response models. The Annals of Statistics, 5, 814-841.

Lawley, D. N. (1943). On problems connected with item selection and test construction. Proceedings, Royal Society of Edinburgh, 61A, 273-287.

Lord, F. M. (1975). Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters. (Research Bulletin 73-33). Princeton, NJ: Educational Testing Service.

Mislevy, R. L., & Bock, R. D. (1984). Item analysis and test scoring with binary logistic models. Mooresville, IN: Scientific Softward, Inc.

Richardson, M. W. (1936). The relationship between difficulty and the differential validity of a test. Psychometrika, 1, 33-49.

Swaminathan, H. (1983). Parameter estimation in item response models. In D. K. Hambleton (Ed.), Applications of Item Response Theory. Vancouver, B. C.: Educational Research Institute of British Columbia.

Swaminathan, H., & Gifford, J. A. (1983). Estimation of parameters in the three-parameter latent trait model. In D. Weiss (Ed.), New Horizons in Testing. New York: Academic Press, 1983.

Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. Journal of Educational Statistics, 7, 175-191.

Swaminathan, H., & Gifford, J. A. (1985). Bayesian estimation in the two-parameter logistic model. Psychometrika, 50, 349-364.

Urry, V. W. (1974). Approximations to item parameters of mental test models and their uses. Educational and Psychological Measurement, 34, 253-269.

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). Logist user's guide. Princeton, NJ: Educational Testing Service.

Wright, B. D., Mead, R., & Bell, S. (1979). BICAL: Calibrating items with the Rasch model. (Research Memorandum 23b). Chicago: Statistical Laboratory, Department of Education, University of Chicago.

Table 1

Frequency Distributions for Raw Scores, Classical Item Difficulties, and Item-Total Biserials

Score Interval	Frequency	Item Difficulty Interval	Frequency	Biserial Interval	Frequency
11-15	2	<.65	1	.200-.299	5
16-18	14	.650-.699	3	.300-.399	5
19-22	19	.700-.749	2	.400-.499	8
23-24	37	.750-.799	4	.500-.599	8
25-26	38	.800-.849	9	.600-.699	9
27-28	43	.850-.899	5	.700-.800	4
29	54	.900-.949	7		
30	45	.950-.999	8		
31	51				
32	56				
33	69				
34	87				
35	84				
36	115				
37	120				
38	107				
39	60				

Note: N=1000 examinees, n=39 items

Table 2

Descriptive Statistics for MML Estimates Based on Three Ability Distributions: Three Parameter Model

Parameter	Ability Distribution	N	E	U	Mean	Standard Deviation
a_g	Normal	1.00	.99	.92	1.83	.98
	Empirical		1.00	.89	1.84	1.07
	Uniform			1.00	1.67	1.10
b_g	Normal	1.00	.98	.95	-1.25	1.05
	Empirical		1.00	.91	[illegible]	1.21
	Uniform			1.00	-1.36	1.10
c_g	Normal	1.00	.96	.70	.24	.21
	Empirical		1.00	.67	.24	.21
	Uniform			1.00	.18	.20

Note: N=250 examinees

Table 3

Descriptive Statistics for BE Estimates Based on Three Ability Distributions: Three Parameter Model

Parameter	Ability Distribution	N	E	U	Mean	Standard Deviation
a_g	Normal	1.00	.99	.98	1.44	.51
	Empirical		1.00	.96	1.41	.52
	Uniform			1.00	1.44	.51
b_g	Normal	1.00	.99	.99	-1.45	1.18
	Empirical		1.00	.99	-1.53	1.27
	Uniform			1.00	-1.32	1.11
c_g	Normal	1.00	.98	.83	.25	.03
	Empirical		1.00	.67	.26	.04
	Uniform			1.00	.25	.02

Note: N=250 examinees

Table 4

Descriptive Statistics for MML Estimates Based on Three Ability Distributions: Three Parameter Common c Model

Parameter	Ability Distribution	N	E	U	Mean	Standard Deviation
a_g	Normal	1.00	.99	.97	1.51	.55
	Empirical		1.00	.96	1.49	.55
	Uniform			1.00	1.49	.57
b_g	Normal	1.00	.99	.99	-1.51	1.16
	Empirical		1.00	.99	-1.58	1.23
	Uniform			1.00	-1.47	1.17
c	Normal	-	-	-	.21	-
	Empirical		-	-	.21	-
	Uniform			-	.19	-

Note: N=250 examinees

Table 5

Descriptive Statistics for MML Estimates Based on Three Ability Distributions: Three Parameter Model

Parameter	Ability Distribution	N	E	U	Mean	Standard Deviation
a_g	Normal	1.00	.99	.88	1.48	.57
	Empirical		1.00	.83	1.48	.58
	Uniform			1.00	1.52	.67
b_g	Normal	1.00	.99	.96	-1.63	1.39
	Empirical		1.00	.97	-1.64	1.39
	Uniform			1.00	-1.74	1.28
c_g	Normal	1.00	.94	.73	.26	.21
	Empirical		1.00	.72	.28	.20
	Uniform			1.00	.17	.20

Note: N=1000 examinees

Table 6

Descriptive Statistics for MML, MML-EAD, BE, and BE-UIP Estimates: Three Parameter Model

Parameter	Estimation Procedure	MML	MML-EAD	BE	BE-UIP	Mean	Standard Deviation
a_g	MML	1.00	.93	.79	.80	1.83	.91
	MML-EAD		1.00	.71	.71	1.78	.87
	BE			1.00	.99	1.53	.53
	BE-UIP				1.00	1.52	.56
b_g	MML	1.00	.97	.93	.92	-1.25	1.05
	MML-EAD		1.00	.93	.92	-1.31	1.19
	BE			1.00	1.00	-1.45	1.18
	BE-UIP				1.00	-1.42	1.08
c_g	MML	1.00	.93	.73	.74	.24	.21
	MML-EAD		1.00	.73	.74	.24	.21
	BE			1.00	.99	.25	.03
	BE-UIP				1.00	.25	.03

Note: N=250 examinees, normal ability distribution

Table 7

Descriptive Statistics for MML, MML-EAD, BE and BE-UIP Estimates: Three Parameter Common c Model

Parameter	Estimation Procedure	MML	MML-EAD	BE	BE-UIP	Mean	Standard Deviation
a_g	MML	1.00	.99	.99	.99	1.52	.56
	MML-EAD		1.00	1.00	.99	1.51	.55
	BE			1.00	.99	1.42	.51
	BE-UIP				1.00	1.49	.52
b_g	MML	1.00	.99	.99	.99	-1.51	1.19
	MML-EAD		1.00	.99	.99	-1.54	1.18
	BE			1.00	.99	-1.49	1.05
	BE-UIP				1.00	-1.47	1.09
c	MML	-	-	-	-	.21	-
	MML-EAD		-	-	-	.22	-
	BE				-	.23	-
	BE-UIP				-	.21	-

Note: N=250 examinees, normal ability distribution

Table 8

Descriptive Statistics for a_g Estimates Obtained Using Various Sample-Size-Estimation-Procedure Combinations: Three Parameter Model

Sample Size	Estimation Method	MML	MML-EAD	BE	BE-UIP	Mean	Standard Deviation
250	MML	1.00	.93	.79	.80	1.83	.91
	MML-EAD		1.00	.71	.71	1.78	.87
	BE			1.00	.99	1.44	.51
	BE-UIP				1.00	1.53	.53
500	MML	1.00	.99	.95	.95	1.35	.59
	MML-EAD		1.00	.93	.93	1.34	.61
	BE			1.00	.99	1.25	.51
	BE-UIP				1.00	1.26	.52
750	MML	1.00	.98	.91	.91	1.55	.52
	MML-EAD		1.00	.91	.91	1.51	.52
	BE			1.00	.99	1.41	.48
	BE-UIP				1.00	1.41	.47
1000	MML	1.00	.99	.94	.95	1.47	.57
	MML-EAD		1.00	.93	.93	1.45	.58
	BE			1.00	.99	1.34	.48
	BE-UIP				1.00	1.36	.49

Note: Normal ability distribution

Table 9

Minimum and Maximum Item Parameter Estimates

		Parameter					
		a_g		b_g		c_g	
Sample Size	Estimation Procedure	Min	Max	Min	Max	Min	Max
250	MML	.5	5.3	- 4.2	.2	.00	.50
	BE	.5	2.7	- 4.3	.7	.18	.34
500	MML	.2	2.8	-12.1	.6	.00	.50
	BE	.3	2.3	- 7.1	.7	.18	.35
750	MML	.4	2.5	- 8.3	.9	.00	.50
	BE	.4	2.3	- 6.7	.9	.15	.44
1000	MML	.5	3.1	- 5.5	.9	.00	.50
	BE	.5	2.2	- 5.6	.9	.15	.42

Note: Normal ability distribution

Table 10

Descriptive Statistics for MML Estimates Based on Incorrect and Fractionally Correct Scoring of Omits: Three Parameter Model

Parameter	Sample Size	Scoring of Omits	W	FC	Mean	Standard Deviation
a_g	250	W	1.00	.97	1.83	.91
		FC		1.00	1.76	.87
	500	W	1.00	.99	1.35	.59
		FC		1.00	1.34	.59
	750	W	1.00	.99	1.56	.52
		FC		1.00	1.55	.52
	1000	W	1.00	.99	1.47	.57
		FC		1.00	1.47	.57
b_g	250	W	1.00	.96	-1.25	1.05
		FC		1.00	-1.29	1.05
	500	W	1.00	.99	-2.18	2.39
		FC		1.00	-2.18	2.33
	750	W	1.00	1.00	-1.56	1.42
		FC		.99	-1.57	1.48
	1000	W	1.00	1.00	-1.63	1.39
		FC		.99	-1.63	1.37
c_g	250	W	1.00	.84	.24	.21
		FC		1.00	.24	.21
	500	W	1.00	.99	.19	.19
		FC		1.00	.19	.20
	750	W	1.00	.99	.18	.19
		FC		1.00	.18	.20
	1000	W	1.00	.99	.26	.21
		FC		1.00	.26	.21

Note: Normal ability distribution

Table 11
Ability Estimate Intercorrelations

MML Item Parameter Estimation						BE Item Parameter Estimation					
ML		EAP		MAP		ML		EAP		MAP	
NBW	BW	NBW	BW	NBW	BW	NBW	BW	NBW	BW	NBW	BW
1.00	.94	.98	.95	.98	.94	.99	.92	.97	.94	.97	.94
	1.00	.94	.96	.94	.97	.95	.99	.94	.95	.95	.96
		1.00	.98	.99	.97	.99	.93	.99	.98	.99	.97
			1.00	.98	.99	.97	.96	.98	.99	.98	.99
				1.00	.98	.99	.94	.99	.98	.99	.97
					1.00	.97	.97	.98	.99	.98	.99
						1.00	.95	.99	.97	.99	.97
							1.00	.95	.97	.95	.97
								1.00	.99	.99	.98
									1.00	.99	.99
										1.00	.99
											1.00

Note: Ability estimates based on item parameter estimates obtained using a sample size of 250 and a normal ability distribution. The sample size for the correlations is also 250.

Table 12

Means and Standard Deviations for Ability Estimates

Estimation Procedure		NBW		BW	
Item	Ability	Mean	Standard Deviation	Mean	Standard Deviation
MML	ML	.05	1.13	.08	1.15
	EAP	-.02	.91	-.07	.85
	MAP	-.07	.86	-.09	.79
BE	ML	.08	1.18	.11	1.20
	EAP	.03	.96	-.03	.89
	MAP	-.01	.91	-.06	.84

Note: Ability estimates based on item parameter estimates obtained using a normal ability distribution

Table 13

Minimum and Maximum Ability Estimates

Estimation Procedure		NBW		BW	
Item	Ability	Minimum	Maximum	Minimum	Maximum
MML	ML	-4.00	2.48	-4.00	3.10
	EAP	-3.99	1.50	-3.94	1.17
	MAP	-4.00	1.35	-3.36	1.03
BE	ML	-4.00	2.45	-4.00	3.32
	EAP	-3.56	1.58	-3.41	1.26
	MAP	-3.55	1.45	-3.41	1.12

Note: Ability estimates based on item parameter estimates obtained using a normal ability distribution.