DOCUMENT RESUME

ED 463 304 TM 033 731

AUTHOR	Li, Yuan H.; Yang, Yu N.
TITLE	The Consistency between the Empirical and the Analytical Standard Errors of Multidimensional IRT Item Estimates.
PUB DATE	2001-04-00
NOTE	32p.; Paper presented at the Annual Meeting of the American Educational Research Association (Seattle, WA, April 10-14, 2001).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Difficulty Level; *Error of Measurement; *Estimation (Mathematics); *Item Response Theory; Maximum Likelihood Statistics
IDENTIFIERS	*Multidimensionality (Tests)

ABSTRACT

An evaluation of the variation of item estimates was conducted for the multidimensional extension of the logistic item response theory (MIRT) model. The empirically determined standard errors (SEs) of marginal maximum likelihood estimation (MMLE)/Bayesian item estimates from 40 items from the ACT Assessment (Form 24b, 1985) were obtained when the same set of items was repeatedly eliminated from test data. These empirically determined SEs were then compared with their corresponding analytical (or formula-based) ones. Both approaches, in general, resulted in similar SE estimates for the same set of items. This empirical comparison implies that the analytical approach has the potential of being used for approximately estimating the magnitudes of SEs of the MMLE/Bayesian item estimates. Tabulation of the analytical SEs for several combinations of item parameters (e.g., low item difficulty, high item discrimination, and low item discrimination was provided as a reference. In addition, the graphical three-dimensional representation of the SEs of item estimates as the bivariate function of item difficulty together with item discrimination was displayed. How to apply the analytical SEs of MIRT item estimates in a MIRT item linking study is illustrated. (Contains 4 tables, 2 figures, and 19 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 463 304

The Consistency between the Empirical and the Analytical Standard Errors of Multidimensional IRT Item Estimates

by

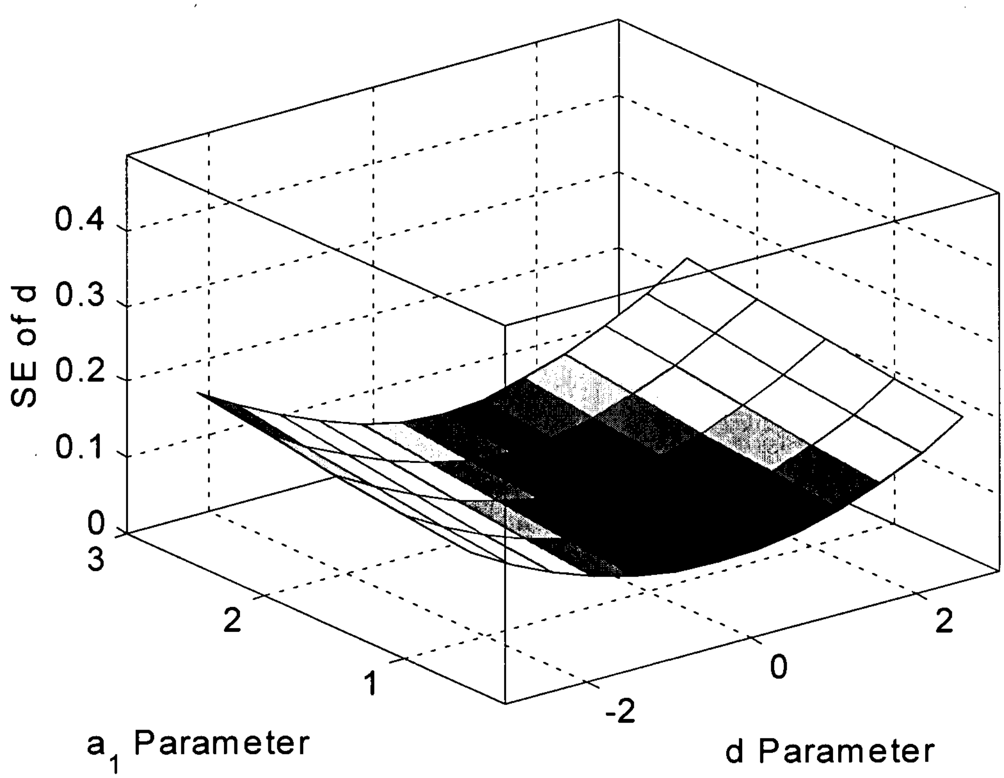
Yuan H. Li
Prince Georges County Public School, Maryland

Yu N. Yang
University of Maryland at College Park

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Y. H. Li
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1



TM033731

*: Paper has been presented at the annual meeting of the American Educational Research Association, April, 10-14, 2001, Seattle, WA

BEST COPY AVAILABLE

The Consistency between the Empirical and the Analytical Standard Errors of Multidimensional IRT Item Estimates

ABSTRACT

An evaluation of the variation of item estimates was conducted for the multidimensional extension of the logistic IRT (MIRT) model. The empirically determined SEs of MMLE (marginal maximum likelihood estimation)/Bayesian item estimates from the forty items (ACT-Form 24b, 1985) were obtained when the same set of items is repeatedly estimated from test data. They seemed to be reasonably small (all less .2) and ready to be used in real testing programs. These empirically determined SEs were then compared with their corresponding analytical (or formula-based) ones. Both approaches, in general, resulted in similar SE estimates for the same set of items. This empirical comparison implies that the analytical approach has the potential of being used for approximately estimating the magnitudes of SEs of the MMLE/Bayesian item estimates.

Tabulation of the analytical SEs for several combinations of item parameters (e.g., low d, high a_1 and low a_2) was provided as a reference. In addition, the graphically 3-D presentation to the SEs of item estimates as the bivariate function of item difficulty together with item discrimination were displayed. Finally, an example of how to apply the analytical SEs of MIRT item estimates on a MIRT item linking study was illustrated.

Key Words: Standard Errors, Parameter Estimates, Multidimensional Item Response Theory

ERIC

I. Introduction

A. Background

When a test is administered to a group of examinees, the interaction of a sample of examinees with a set of test tasks might result in test data that appear to be unidimensional in some instances but multidimensional in other instances (Ackerman, 1992) because this set of test items can be sensitive to several traits. Furthermore, this group of examinees may vary in several latent abilities (Ackerman, 1992). A presumed single trait dimension for test data that are actually multidimensional might jeopardize the invariant feature of item response theory (IRT) models (Ackerman, 1994; Reckase, 1985). The results from the Li and Lissitz's study (2000a) suggested that multidimensional IRT (MIRT) models can be applied to not only multidimensional data but also to unidimensional test data as well. The fit of MIRT MIRT models to unidimensional data will generate item discrimination estimates (or factor loadings) that approach zero for the overidentified dimensions but does little harm in terms of the IRT invariant feature. In order to avoid obtaining unbiased parameter estimates, it seems apparent that overestimating the number of dimensions of a set of test data would be a better choice than underestimating the number of dimensions (Reckase & Hirsch, 1991).

MIRT modeling takes advantage of more flexibility of fitting test data than unidimentional models, but requires that more model parameters be estimated. This latter factor might result in parameters that are less accurate and stable when sample sizes are not large enough. The practical utility of MIRT models relies on the capability of obtaining reasonably accurate item estimates (Miller, 1991). The magnitude of a standard error (SE) of an item parameter is used to measure the precision of an item estimate and will become a critical criterion to gauge MIRT's feasibility for future practical uses.

The estimated SE for an item estimate is strongly associated with the parameter estimation method. The joint maximum likelihood method (JMLE) was one of the estimation methods. During the process of JMLE, the asymptotic variances and covariances of MIRT item estimates can be approximately estimated by inverting the associated information matrix for the item estimates at the last iteration when we treat ability estimates as true values (Carlson, 1987). The diagonal of the inverse of the information matrix contains the corresponding error variances of item estimates. And the square roots of these elements in diagonal are the approximate asymptotic SEs of item estimates. The MIRT program (Carlson, 1987), implementing the JMLE estimation method, has options for users to obtain this type of information. One estimation problem for the JMLE item estimates is that they are not statistically consistent as the number of examinees increases (Baker, 1992). That is why this approach has become less widely used than the MMLE (marginal maximum likelihood estimation) /Bayesian approach that generally produces better estimates with small sample sizes. The MMLE/Bayesian estimation involves the incorporation of the additional information of the priors of item estimates into the MMLE likelihood function.

The MMLE /Bayesian approach similar to the JMLE procedure is capable of approximating the asymptotic SEs of MIRT item estimates when the distribution of examinee abilities is exactly specified in the likelihood function. As a matter of fact, the published and accessible MIRT software, TESTFACT (Wilson, Wood & Gibbons, 1991) using the MMLE/Bayesian, does not provide this type of information. Instead, the analytical approach (or formula-based, Thissen & Wainer, 1982, will be introduced later) fills this gap by predicting SE's values without real test data when a set of MIRT item estimates (e.g., yielded from TESTFACT) are given. The fundamental assumption used for deriving the formula for computing the analytical SEs is that item parameters are estimated by the maximum likelihood (ML) rather than the MMLE/Bayesian approach. Does this assumption have any significant impact on approximating the SEs of MMLE/Bayesian item estimates? The accuracy of the SE estimates for unidimenstional IRT items through the analytical approach has been examined by the Li and Lissitz's study (2000b). Their study demonstrated the analytical approach is suitable to approximate the SE estimates for the two-parameter model and the generalized partial credit model (Muraki, 1992). Their findings encourage test practitioners to further explore the possibility of using the analytical approach for predicting the SEs of item estimates for the MIRT models as well.

Another method, the least squares approach implemented in NOHARM (1988, Fraser & McDonald), has been used in several studies (e.g., Miller, 1991; Reckase, 1985) for estimating MIRT parameter estimates, but the least squares approach is not directly available to approximate the SEs of MIRT item estimates as do the JMLE and MMLE/Bayesian. Miller (1991) attempted the empirically determined approach by repeated samplings to obtain the SEs of the least-squares based MIRT estimates. In Miller's study (1991), a population sample of 30,000 examinees were drawn at random from 140,000 cases. Ten replication samples of n=2000 each were then drawn at random, with replacement from the presumed population. The average of the empirical SEs for item difficulty (d) was 0.15, and for the first (a_1), second (a_2) and third (a_3) discrimination parameters were 0.12, 0.14 and 0.15, respectively. The results from Miller's study provided test practitioners with valuable information about how large the empirical SEs of item parameter estimates from a real dataset could be, although the magnitudes of SE estimates from that study could be unstable due to the small number of replications in that study.

ERIC

B. Research Purposes

MIRT modes have not currently been employed in real testing programs. It is so essential that test practitioners know how reliable a set of MIRT item estimates are before employing them in practical testing situations. The empirically determined approach is very tedious for obtaining the SEs of parameter estimates, but it will produce rather stable and accurate SE estimates for a set of item estimates as the number of replications increases. Accordingly, the empirically determined approach was adopted in this study for serving two purposes. One is that the empirically determined SE results would provide test practitioners with a sense of how large the SEs of MIRT parameter estimates might be when the MMLE/Bayesian approach is applied. The other is that they will be used as a comparison base with those obtained from the analytical approach.

The level of consistency of SEs yielded from both the empirically determined and the analytical methods was used to evaluate the feasibility of using the analytical method for predicting the SEs of MIRT item estimates. If the level of consistency between two measures for SE estimates is relatively high, the analytical-based SEs of MIRT item estimates would be tabulated under some common testing situations for reference purposes as has been done by Thissen & Wainer (1982) for unidimentional IRT item estimates.

Without real test data the three-D graphical presentation for analytical SEs of item estimates has been used for detecting the possible problems of applying the ML estimation method to the Three-PL model (Thissen & Wainer, 1982) and to the GPCM model (Li & Lissitz, 2000b). As the MIRT model (Reckase,1985) has been increasingly used in research studies, the extension of this graphical procedure to the MIRT model will provide test practitioners with a better understanding of this model. The three-D graphical presentation for the two-dimensional case is included in this study.

II. Methods for Approximating the SEs of MIRT Item Estimates

A. Multidimensional Logistic IRT Models

The model illustrated below is a multidimensional extension of the three-parameter logistic model (M3PL). This model hypothesizes that the probability of a correct response, $u_{ij}=1$, by person j to item i, given an individual's m-dimensional latent abilities, $\underline{\theta}_j$, is (refer to Reckase, 1985): $P(u_{ij}=1 \mid \underline{a}_i, d_i, c_i, \underline{\theta}_j) = c_i + (1-c_i)\frac{e^{Z_{ij}}}{1+e^{Z_{ij}}}$ (1)

where,

$$Z_{ij} = D\left(\sum_{k=1}^{m} a_{ik}\theta_{jk}\right) + d_i \quad (2)$$

$\underline{a}_i$ is a m-dimensional vector of item discrimination parameters,

d_i is a location parameter related to item difficulty

c_i is a pseudo-guessing parameter and

D is a scaling constant (1.702).

The scaling factor D is included in the model to make the logistic function as close as possible to the normal ogive function (Baker, 1992). Since the terms in Equation 2 are additive, being low on one latent trait can be compensated for by being high on the other latent traits. Thus, this model is called a compensatory model (Reckase, 1985) because the terms are additive in the logit. A multidimensional extension of the two-parameter logistic model

ERIC

(M2PL) is obtained if the guessing parameter c_i is constrained to zero for all items in Equation 1 above.

B. The Empirically Determined SEs of MMLE/Bayesian MIRT Item Estimates

The empirically determined SEs of item estimates can be calculated from the real test data as described by Miller (1991) or can be obtained through repeated data generations as illustrated below. When the MMLE/Bayesian estimation approach was used for item estimates, the empirically determined SEs of MMLE/Bayesian MIRT item estimates are obtained in the following manner:

(1). Generate a test dataset by using the known MIRT item parameters and a set of simulees's ability parameters;

(2). Calibrate item parameter estimates, using the MMLE/Bayesian estimation method;

(3). Transform the metric of the estimated parameters to the one defined by the true parameters (for detailed procedures on MIRT item linking, see Li and Lissitz, 2000a);

(4). Repeat steps 1 through 3 numerous times (e.g., 100 times), resulting in a large number of estimates for each individual parameter.

The SE of an item estimate is obtained by computing the standard deviation of the replicated (e.g. 100) item estimates. In addition, the SE of an item estimate can be calculated through the values of BIAS and RMSE, defined below, if both variables are available. The latter method is preferred, if available, because the relationship to BIASs and SEs for item estimates can also be evaluated when needed.

(5). Calculate the BIAS and RMSE (root mean squared error) for each of the parameter estimates by the formulas shown below.

$$\mathrm{BIAS}(H_i) = \frac{\sum_{i=1}^{r}(\hat{H}_i - H_i)}{r} \tag{3}$$

$$\mathrm{RMSE}(H_i) = \sqrt{\frac{\sum_{i=1}^{r}(\hat{H}_i - H_i)^2}{r}} \quad \text{and} \tag{4}$$

where H_i is the true item parameter, $\hat{H}_i$ is the corresponding estimated item parameter, and r is the number of replications, in which r equals 100 in this study.

RMSE is a measure of total error of estimation that consists of the systematic error (BIAS) and random error (SE). These three indices are related to each other as follows (De Ayala & Sava-Bolesta,1999):

$$\mathrm{RMSE}(H_i)^2 \cong \mathrm{SE}(H_i)^2 + \mathrm{BIAS}(H_i)^2 \tag{5}$$

The empirical MMLE/Bayesian SE of an item estimate is approximately calculated by:

$$\mathrm{SE}(H_i) \cong \sqrt{\mathrm{RMSE}(H_i)^2 - \mathrm{BIAS}(H_i)^2} \tag{6}$$

C. The Analytical SEs of MIRT ML Item Estimates

Sample size, the shape of examinees' abilities and the characteristic of test items can cause errors in the parameter estimates (Hambleton, Jones & Rogers, 1993; Stocking, 1990; Thissen & Wainer, 1982). A mathematical expression for this relationship has been developed by Thissen and Wainer (1982) for unidimensionally dichotomous IRT models and was modified for multidimensionally dichotomous IRT models (Li & Lissitz, 2000a). The procedures illustrated below were used in this study for predicting the SEs of MIRT item estimates.

For an item i, the likelihood of the observed responses for N independent examinees is:

$$L = \prod_{j=1}^{N} P_j^{u}(1 - P_j)^{1-u} \tag{7}$$

where P can be calculated from a M2PL model, u=1 for correct response; u=0 for incorrect response. The log likelihood of Equation 7 is

$$\log L = \sum_{j=1}^{N}\left[u \log(P_j) + (1-u)\log(1-P_j)\right] \tag{8}$$

The maximum likelihood estimates of each parameter ($\underline{a}_i$, d_i,) are located where the partial derivatives of Equation 8 are zero. Let ξ represents the M2PL item parameters ($\underline{a}_i$, d_i,). Given a density of θ (e.g. multivariate Gaussian with MVN(**0**, **I**)), for any parameter ξ_s and ξ_t, the negative expected value of the second derivative of the log likelihood function in Equation 8 has the form (refer to Thissen, Wainer, 1982),

$$-E\left(\frac{\partial^2 \log L}{\partial\xi_s \partial\xi_t}\right) = N\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\left\{\left(\frac{1}{PQ}\right)\left(\frac{\partial P(\theta)}{\partial\xi_s}\frac{\partial P(\theta)}{\partial\xi_t}\right)\right\}\Phi_i(\theta)\,d\theta_1\cdots d\theta_m \tag{9}$$

where E is the expectation and Q=1-P. Equation 9 requires the derivatives of P(θ) with respect to its parameters. These derivatives of P(θ) can be substituted in Equation 9 to give a 3 x 3 (for the M2PL model) information matrix corresponding to the triplet item parameters (d, a_1 and a_2). The inverse of that information matrix is the asymptotic variance-covariance matrix of the three parameters. The square roots of the diagonal elements of the variance-covariance matrix are the asymptotic standard errors of the parameters.

ERIC

The numerical approximation of the multiple integral in Equation 9 can be calculated by the multiple-dimensional Gauss-Hermite quadrature (Baker, 1992). Equation 10 is presented for the two-dimensional case,

$$N\sum_{q_2=1}^{q}\sum_{q_1=1}^{q}\left\{\left(\frac{1}{PQ}\right)\left(\frac{\partial P(X)}{\partial \xi_s}\frac{\partial P(X)}{\partial \xi_t}\right)\right\}A(X_{q_1})A(X_{q_2}) \tag{10}$$

where X is a quadrature point in one of two ability dimensions, q is the number of quadratures in this ability dimension and A(X) is the corresponding weight of the quadrature. The number of quadrature points for numerical integration is set at forty for each dimension in this study.

To summarize, the analytical SEs of a set of ML item estimates for an item are a function of the IRT model, the sample size and the shape of the examinees' abilities.

III. Methodology

A. The Empirically determined SEs of MIRT MMLE/Bayesian Item Estimates

The M2PL model (Reckase, 1985) was used in this study. MIRT item estimates for the 40 items were from ACT Form 24B (Reckase, 1985). The RMSEs and BIASs of MMLE/Bayesian item estimates were calculated when the 40 items were repeatedly calibrated by TESTFACT from 100 simulation test data (for details, see Li's study, 1997; Li & Lissitz, 2000a). The empirically determined SEs of MIRT item estimates were calculated by Equation 6, using the RMSE and BIAS information.

B. The Analytical SEs of MIRT ML Item Estimates

The same 40 sets of item estimates were also used to calculate the analytical SEs of item estimates. The weights (A(Xs)) for all quadrature points used for the analytical SE estimates came from the estimated posterior distribution of abilities reported from the TESTFACT output when the item parameters were estimated by the MMLE/Bayesian approach. The same sample size, 2000, used to generate item response data in Li and Lissitz's study (2000a), was used here.

C. Data Analysis

Descriptive statistics of the SE Index of item parameter estimates for the analytical and the empirically determined data were calculated. A t-test for dependent observations was then conducted to compare the impact of the estimation method on the precision of SE estimates for item parameters. The dependent t-test was chosen because the values of SEs for the same set of item parameters were repeatedly calculated by the two methods so that the Log[SE] (the log transformation of SE, refers to Harwell, Stone, Hsu & Kirisci, 1996) of each of various item parameter estimates was treated as a repeated- measure across two methods.

The Pearson correlation coefficient between two measures, across test items, was calculated for each of the various item estimates. The plots of SEs of item estimates as a function of true item parameters for these two approaches were graphed.

IV. Results and Discussions:

A. The SEs of MIRT Item Estimates

The SEs of the set of item parameters (from ACT Form-24B, Reckase, 1985) were calculated by the empirically determined and analytical approaches with a sample size 2000 and are tabulated in Table 1.

ERIC

In Table 1, the values of the second column are d-parameters, the next two columns used for showing their corresponding SEs separately calculated by the empirically determined (labeled EMB) and the analytical (labeled ANA) approaches, respectively. The values of a_1-parameters are in the fifth column and their corresponding SEs computed from the two approaches are presented in the next two columns. The values of the eighth column are a_2-parameters, and the next two columns are used to show their corresponding SEs for the two approaches. For example, Item 2 has MIRT parameters of $d=0.17$, $a_1=1.22$ and $a_2=0.02$; the corresponding empirically determined SEs for this set of item parameters are 0.032, 0.034 and 0.047. They are quite similar to those calculated from the analytical approach that resulted in 0.055, 0.040 and 0.036 for the same item. The average value of each parameter and its corresponding SE is presented in the last row of Table 1.

The magnitudes of SEs for the 40 sets of item estimates in Table 1 from the empirically determined approach were the empirical SEs of MIRT MMLE/Bayesian item estimates and ,in general, are rather small (less than .2). The question of whether the complex MIRT models have the capability of obtaining reasonably accurate item estimates might have been addressed based on this result. MIRT models, essentially, have more slope (discrimination) parameters to be estimated than the unidimentional IRT models do and this may cause greater variation of item estimates. The risk of obtaining unreliable MIRT item estimates did not occur in the data examined in this study. The theoretically sound estimator of MMLE/Bayesian and an appropriate method used for item metric conversion (or MIRT item linking, see Li and Lissitz, 2000a) could be two key factors affecting this result.

ERIC

Table 1:
The SEs of MIRT Item Estimates from ACT-Form 24b (Reckase, 1985) by the Empirically Determined and Analytical Approaches (N = 2000, Replications =100, for the Empirically determined Approach)

Item #	d	SE of d		a_1	SE of a_1		a_2	SE of a_2	
		EMB	ANA		EMB	ANA		EMB	ANA
01	0.170	0.037	0.058	1.220	0.075	0.055	0.020	0.057	0.034
02	0.440	0.032	0.055	0.710	0.034	0.040	0.530	0.047	0.036
03	0.440	0.044	0.067	1.720	0.094	0.077	0.180	0.072	0.039
04	0.690	0.035	0.064	1.330	0.066	0.061	0.340	0.049	0.038
05	0.380	0.045	0.071	2.000	0.120	0.091	0.000	0.079	0.041
06	0.910	0.047	0.081	2.000	0.091	0.095	0.980	0.064	0.060
07	0.540	0.038	0.061	1.220	0.074	0.056	0.140	0.047	0.035
08	-0.210	0.043	0.067	1.350	0.056	0.066	1.150	0.054	0.059
09	0.120	0.039	0.068	1.920	0.111	0.087	0.000	0.081	0.040
10	-0.280	0.035	0.058	1.200	0.068	0.055	0.120	0.052	0.034
11	0.020	0.042	0.074	1.540	0.090	0.079	1.790	0.106	0.088
12	-0.830	0.039	0.070	1.530	0.076	0.070	0.480	0.052	0.042
13	-0.490	0.031	0.054	0.510	0.040	0.036	0.650	0.050	0.038
14	-0.680	0.029	0.054	0.690	0.043	0.038	0.190	0.038	0.031
15	-1.080	0.045	0.071	0.680	0.060	0.045	1.210	0.076	0.058
16	-1.000	0.048	0.068	0.510	0.063	0.041	1.210	0.092	0.057
17	-1.920	0.070	0.102	0.010	0.123	0.043	1.940	0.174	0.093
18	-1.360	0.048	0.074	0.760	0.049	0.046	0.990	0.078	0.052
19	-0.990	0.042	0.065	0.290	0.054	0.036	1.100	0.090	0.053
20	-1.610	0.047	0.073	0.420	0.045	0.039	0.750	0.055	0.045
21	1.460	0.045	0.090	1.810	0.073	0.087	0.860	0.061	0.056
22	0.670	0.043	0.068	1.570	0.080	0.071	0.360	0.051	0.040
23	0.100	0.034	0.053	0.860	0.050	0.043	0.190	0.043	0.032
24	0.380	0.046	0.069	1.860	0.102	0.084	0.290	0.076	0.041
25	0.170	0.050	0.069	1.190	0.086	0.063	1.570	0.092	0.075
26	0.030	0.034	0.053	0.870	0.051	0.043	0.000	0.046	0.031
27	-0.490	0.040	0.062	1.000	0.055	0.051	0.890	0.061	0.048
28	0.290	0.037	0.061	1.270	0.069	0.058	0.470	0.047	0.039
29	0.080	0.036	0.057	1.060	0.054	0.050	0.450	0.041	0.037
30	-0.300	0.033	0.055	0.960	0.053	0.046	0.220	0.036	0.033
31	-0.210	0.037	0.061	1.410	0.074	0.063	0.040	0.060	0.036
32	-0.690	0.033	0.053	0.540	0.043	0.035	0.230	0.039	0.031
33	-0.560	0.034	0.056	0.720	0.045	0.040	0.550	0.043	0.037
34	-0.380	0.050	0.076	1.660	0.099	0.084	1.720	0.093	0.086
35	-0.910	0.049	0.069	0.880	0.057	0.049	1.120	0.075	0.056
36	-0.950	0.052	0.065	0.240	0.063	0.036	1.140	0.095	0.054
37	-0.960	0.034	0.062	0.760	0.045	0.043	0.590	0.043	0.039
38	-1.570	0.084	0.089	0.390	0.079	0.044	1.770	0.149	0.083
39	-0.810	0.049	0.063	0.490	0.060	0.039	1.100	0.087	0.053
40	-1.560	0.055	0.076	0.480	0.048	0.041	1.000	0.077	0.052
Mean	-0.324	0.043	0.067	1.041	0.068	0.056	0.708	0.068	0.048

ERIC

B. Comparisons Between the Analytical SEs and the Empirically determined SEs

Figure 1 presents plots of the SEs as a function of true parameters (item difficulty, d, together with the item discriminations, a_1 and a_2) for the analytical (labeled ANA) and the empirically determined (labeled EMB) methods. Table 2 shows summary descriptive statistics for the SE, computed across 40 items, for each method.

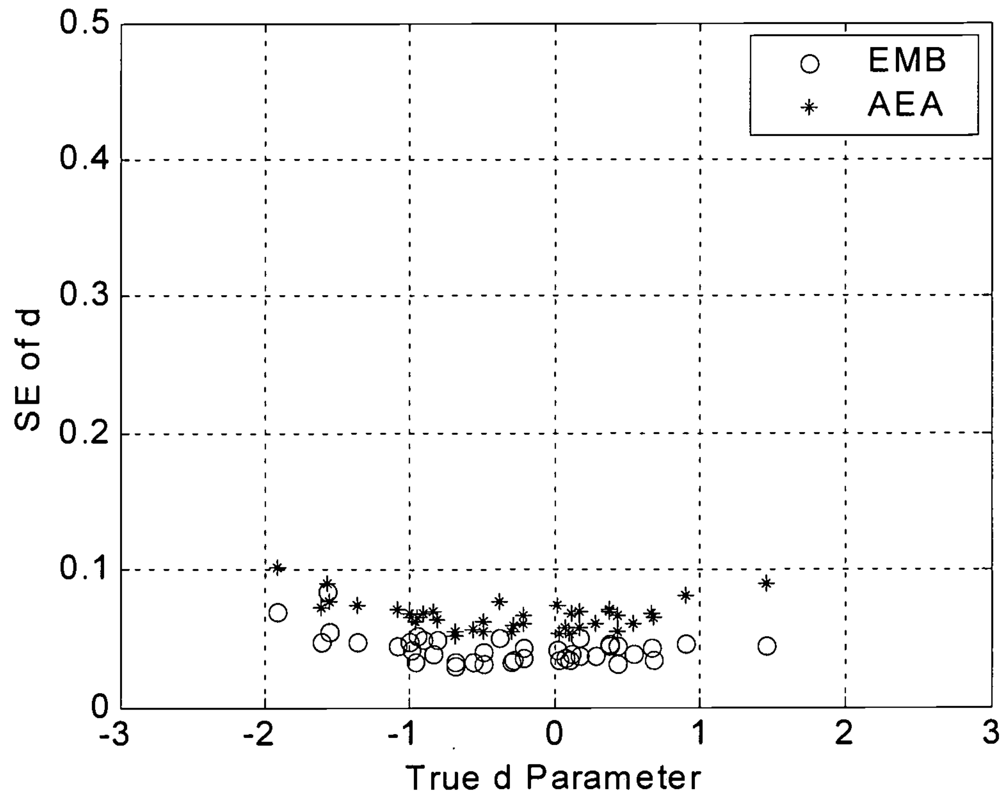


Figure 1a. SE of d as a function of the true d-parameter for the two-dimensional MIRT model.

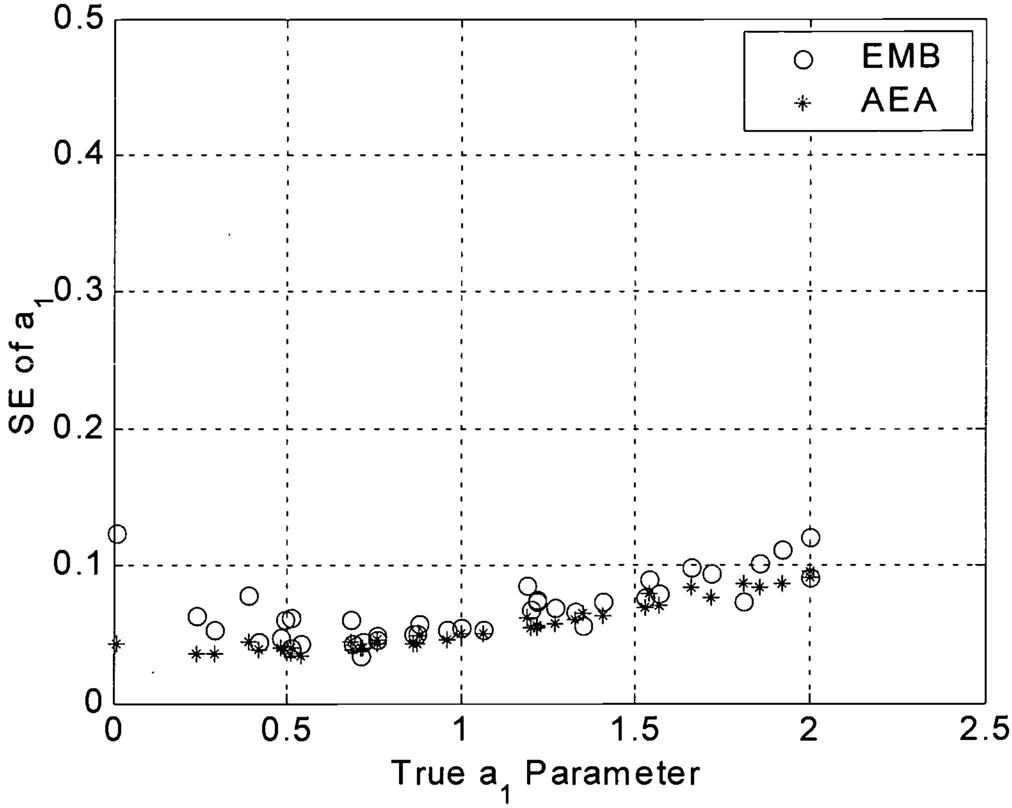


Figure 1b. SE of a_1 as a function of the true a_1-parameter for the two-dimensional MIRT model.

The results from Figure 1 and Table 2 indicated that both methods yielded quite similar SEs of item estimates. The dependent-t statistics (see Table 2) for the log transformation of the parameters (d, a_1, and a_2) showed statistically significant differences between the two methods. It seems likely that there is no practical meaning for these statistically significant differences, since these differences (.03 for d, .01 for a_1, and .02 for a_2) are so small.

The correlation coefficients (Table 2) between analytical and empirically determined measures, across 40 items, were .83, .63 and .79 for the parameters, d, a_1 and a_2, respectively.

To summarize, the magnitudes of the analytical SEs of ML item estimates were highly consistent with those obtained from the empirically determined SEs of MMLE/Bayesian item estimates. This finding implies that although the analytical SEs of MIRT item estimates was derived on the assumption that items are estimated by the ML estimation method, they can be substituted for those from MMLE/Bayesian item estimates when the latter are unavailable, difficult to calculate exactly or needed at some other purposes (discussed later).

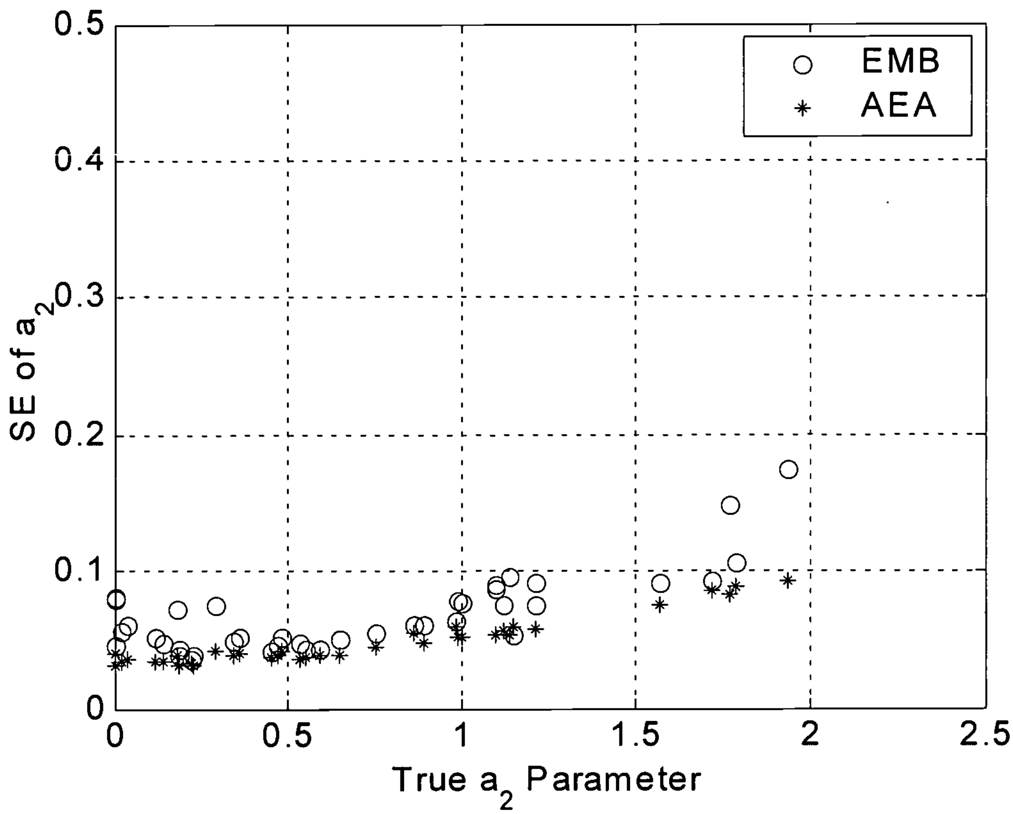


Figure 1c. SE of a_2 as a function of the true a_2-parameter for the two-dimensional MIRT model.

Table 2
Descriptive Statistics of SE Index of MIRT Item Parameter Estimates, dependent t Tests, and Pearson Correlation Coefficients, for the AEA and EMB methods (N=2000, Replications for EMB = 1000)

Method and Parameter	Mean	Number of Items	AEA Mean	AEA Min	AEA Max	EMB Mean	EMB Min	EMB Max	t	r
Two-Dim										
d	-.32	40	.07	.05	.11	.04	.03	.08	25.17***	.83
a1	1.04	40	.06	.04	.10	.07	.03	.12	-5.88***	.63
a2	.71	40	.05	.03	.10	.07	.04	.17	-10.50***	.79

* P < .05; ** P< .01; *** P< .001

C. Tabulate and Graph the Analytical SEs of MIRT Item Estimates

There were few studies related to the evaluation of the magnitude of SEs for MIRT item estimates. This could result in the test practitioners inability to realize how small the SEs of MIRT parameter estimates are that are obtained under a specific condition. Tabulating the analytical SEs by the different combinations of item parameters (e.g., low d, low a_1 with high a_2) is an alternative method to help test practitioners better comprehend the possible values of the estimated SEs of item parameters.

Table 3 provides the SEs of item difficulty (d) as a three-variable (d, a_1 and a_2) function for the two-dimensional M2PL Model with a sample size, N=1000. For example, given a set of MIRT item estimates of d=0, a_1=1 and a_2=1, the SE of d is .073. The SEs in this table can be applied to other sample sizes. For instance, for N=2000, the corresponding SEs will be the current SEs presented in this table multiplied by a constant that equals : $\frac{1}{\sqrt{\frac{2000}{1000}}}$.

Similarly, for N=3000, the constant equals $\frac{1}{\sqrt{\frac{3000}{1000}}}$.

The procedure for computing the constant is (refer to Thissen & Wainer, 1982): first, compute the ratio of the new sample size to 1000, second, take the square root of this ratio, finally, take reciprocal of this ratio.

Table 4 provides the SEs of the first-dimensional item discrimination (a_1) as a three-variable (d, a_1 and a_2) function for the two-dimensional M2PL Model under N=1000. Similarly, the SEs in Table 4 can be used for the SEs of a_2 and be applied to other sample sizes using the same method of calculation.

Table 3:
SE of Item Difficulty (d) as a Three-variable Function of Item Difficulty (d) together with the Two Item discriminations (a_1 and a_2) for the Two-dimensional M2PL Model with N=1000.

Item Discrimination		**Item Difficulty d**												
a_2	**a_1**	**-3.0**	**-2.5**	**-2.0**	**-1.5**	**-1.0**	**-0.5**	**0.0**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**
0.5	**0.5**	.171	.139	.114	.095	.082	.074	.072	.074	.082	.095	.113	.139	.171
	1.0	.183	.151	.126	.106	.092	.083	.080	.083	.092	.106	.126	.151	.183
	1.5	.195	.164	.139	.118	.103	.093	.089	.093	.103	.118	.139	.165	.196
	2.0	.209	.178	.152	.130	.113	.103	.099	.103	.113	.130	.152	.179	.210
	2.5	.223	.192	.165	.142	.124	.112	.108	.112	.124	.142	.165	.192	.224
1.0	**0.5**	.183	.151	.126	.106	.092	.083	.080	.083	.092	.106	.126	.151	.183
	1.0	.191	.160	.134	.114	.099	.089	.086	.089	.099	.114	.134	.160	.192
	1.5	.202	.171	.145	.124	.108	.098	.094	.098	.108	.124	.145	.171	.203
	2.0	.214	.183	.157	.134	.117	.106	.102	.106	.117	.134	.157	.184	.215
	2.5	.228	.196	.168	.145	.127	.115	.111	.115	.127	.145	.169	.196	.228
1.5	**0.5**	.195	.164	.139	.118	.103	.093	.089	.093	.103	.118	.139	.165	.196
	1.0	.202	.171	.145	.124	.108	.098	.094	.098	.108	.124	.145	.171	.203
	1.5	.211	.180	.154	.132	.115	.104	.100	.104	.115	.132	.154	.181	.212
	2.0	.222	.191	.164	.141	.123	.111	.107	.111	.123	.141	.164	.191	.223
	2.5	.234	.202	.174	.150	.131	.119	.115	.119	.131	.150	.174	.203	.235
2.0	**0.5**	.209	.178	.152	.130	.113	.103	.099	.103	.113	.130	.152	.179	.210
	1.0	.214	.183	.157	.134	.117	.106	.102	.106	.117	.134	.157	.184	.215
	1.5	.222	.191	.164	.141	.123	.111	.107	.111	.123	.141	.164	.191	.223
	2.0	.232	.200	.172	.148	.130	.118	.113	.118	.130	.148	.172	.200	.233
	2.5	.243	.210	.181	.157	.137	.124	.120	.124	.137	.157	.181	.210	.243
2.5	**0.5**	.223	.192	.165	.142	.124	.112	.108	.112	.124	.142	.165	.192	.224
	1.0	.228	.196	.168	.145	.127	.115	.111	.115	.127	.145	.169	.196	.228
	1.5	.234	.202	.174	.150	.131	.119	.115	.119	.131	.150	.174	.203	.235
	2.0	.243	.210	.181	.157	.137	.124	.120	.124	.137	.157	.181	.210	.243
	2.5	.252	.219	.189	.164	.143	.130	.125	.130	.143	.164	.189	.219	*

*: Unavailable due to inappropriate combination for a set of item parameters (d=3.0, a_1=2.5 and a_2=2.5).

Table 4:
SE of Item Discrimination (a_1) as a Three-variable Function of Item Difficulty (d) together with the Two Item discriminations (a_1 and a_2) for the Two-dimensional M2PL Model with N=1000.

Item Discrimination		**Item Difficulty d**												
a_2	**a_1**	**-3.0**	**-2.5**	**-2.0**	**-1.5**	**-1.0**	**-0.5**	**0.0**	**0.5**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**
0.5	**0.5**	.078	.068	.060	.054	.051	.049	.048	.049	.051	.055	.061	.069	.079
	1.0	.090	.083	.077	.073	.070	.069	.068	.069	.071	.074	.078	.084	.092
	1.5	.114	.108	.104	.100	.098	.097	.096	.097	.098	.101	.105	.109	.116
	2.0	.146	.141	.137	.134	.131	.130	.130	.130	.132	.134	.137	.142	.147
	2.5	.184	.179	.175	.172	.170	.169	.168	.169	.170	.172	.175	.179	.185
1.0	**0.5**	.073	.066	.061	.058	.055	.054	.053	.054	.055	.058	.062	.067	.074
	1.0	.090	.084	.080	.077	.074	.073	.073	.073	.075	.077	.081	.085	.091
	1.5	.116	.111	.107	.104	.102	.100	.100	.101	.102	.104	.108	.112	.117
	2.0	.148	.144	.140	.137	.135	.134	.133	.134	.135	.137	.140	.145	.150
	2.5	.186	.182	.178	.175	.173	.172	.172	.172	.173	.176	.179	.182	.187
1.5	**0.5**	.073	.068	.065	.062	.060	.059	.059	.059	.060	.062	.065	.069	.073
	1.0	.092	.087	.084	.081	.080	.079	.078	.079	.080	.082	.085	.088	.093
	1.5	.119	.115	.111	.109	.107	.106	.106	.106	.107	.109	.112	.116	.120
	2.0	.152	.148	.145	.142	.140	.139	.139	.139	.141	.143	.145	.149	.153
	2.5	.190	.186	.183	.180	.179	.178	.177	.178	.179	.181	.184	.187	.191
2.0	**0.5**	.075	.072	.069	.067	.066	.065	.065	.065	.066	.067	.069	.072	.075
	1.0	.095	.092	.089	.087	.086	.085	.085	.085	.086	.087	.089	.092	.096
	1.5	.123	.120	.117	.115	.113	.112	.112	.113	.114	.115	.117	.120	.124
	2.0	.157	.153	.151	.148	.147	.146	.146	.146	.147	.149	.151	.154	.158
	2.5	.196	.192	.189	.187	.185	.184	.184	.185	.186	.187	.190	.193	.197
2.5	**0.5**	.078	.076	.074	.072	.071	.071	.071	.071	.072	.073	.074	.076	.079
	1.0	.099	.097	.094	.093	.092	.091	.091	.091	.092	.093	.095	.097	.100
	1.5	.128	.125	.123	.121	.120	.119	.119	.119	.120	.122	.123	.126	.129
	2.0	.163	.160	.157	.155	.154	.153	.153	.154	.154	.156	.158	.161	.164
	2.5	.202	.199	.196	.194	.193	.192	.192	.192	.193	.195	.197	.200	*

*: Unavailable due to inappropriate combination for a set of item parameters (d=3.0, a_1=2.5 and a_2=2.5).

When test practitioners are working with the two-dimensional logistic MIRT model, Tables 3 and 4 are useful references for predicting the SEs of MIRT item estimates. In addition, similar tables like Tables 3 and 4 can be made for different conditions (e.g., for the three-dimensional MIRT models) if needed. Without real test data, we are able to explore the SE's characteristics of item parameters for all commonly used testing conditions using the analytical approach. This is almost impossible when employing the empirically determined method. The SEs of d in Table 3 can be graphically presented in a 3-D display when the second-dimensional discrimination parameter is set at a constant, such as 1 used in Figure 2. Figure 2a is the 3-D plot for the SEs of in Table 3. This presents plots of SEs of item difficulty as the

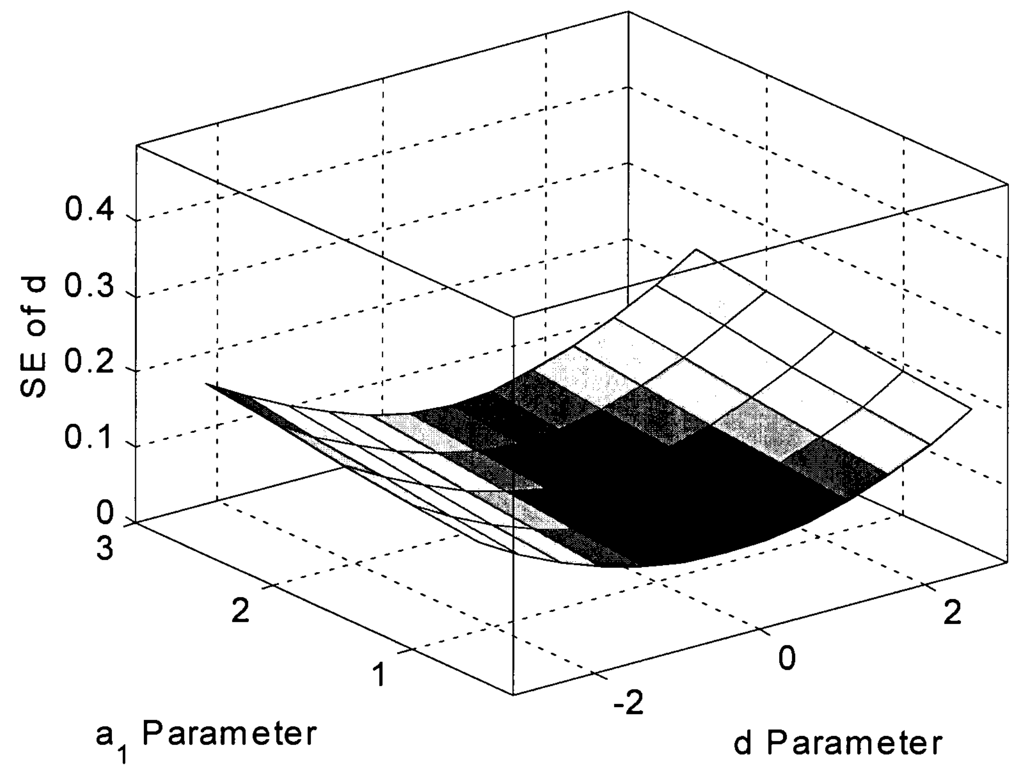


Figure 2a. The SEs of ds shown as the bivariate function of both d and a_1 parameters for the two-dimensional MIRT model when the a_2 is set to 1.

bivariate function of both item difficulty and the first-dimensional discrimination parameters. This diagram demonstrates that an extreme item (hard or easy) is more likely to have more measurement error. Figure 2b turns its focus on the SEs of a_1-parameters, where SEs of a_1

were from Table 4. This plot raises an interesting issue in that the SE of the a_1-parameter is continuing to increase as the a_1-parameters is increasing.

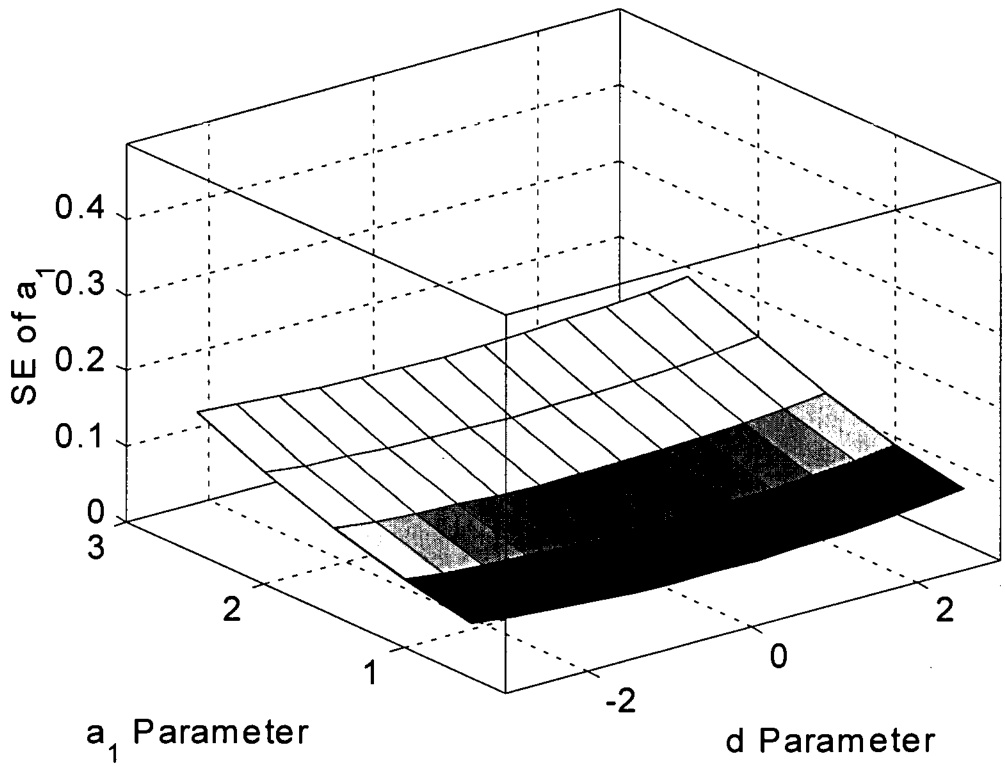


Figure 2b. The SEs of a_1 shown as the bivariate function of both d and a_1 parameters for the two-dimensional MIRT model when the a_2 is set to 1.

D. Application of the Analytical SEs of Item Estimates on MIRT Item-Linking Studies

This section discusses the application of the analytical SEs of item estimates to item-linking studies. Given a set of known (or true) item parameters, the analytical SEs of item estimates can help researchers generate the set of observed item estimates without the process of data generation that is often needed in the simulation studies.

Conceptually, a numerical value of an item parameter estimate can be decomposed into three components: true item parameter value, a random error and bias. When bias is assumed to be zero and a set of true item parameters for an item is given, the procedure of adding "reasonable numerical values as random errors" to this set of true parameters to create a set of observed item estimates is illustrated below.

When the latent trait distribution of 2000 examinees' abilities is multidimenstionally distributed as MVN(**0**,**I**), the variance-covariance matrix **V** shown below for a set of item parameters, d=0.44, a_1=0.71 and a_2=0.53 (refer to Item 2 in Table 1) can be predicted using Equation 10 :

$$
\begin{array}{cc}
 & \begin{array}{ccc} \;\; d & \quad a_1 & \quad a_2 \end{array} \\
V = & \begin{bmatrix} .0030 & .0003 & .0003 \\ .0003 & .0016 & .0004 \\ .0003 & .0004 & .0013 \end{bmatrix}
\end{array}
$$

The square root of the diagonal elements of the matrix **V** are the asymptotic standard errors of the parameters. They are .547, .040 and .036 for the parameters, d, a_1 and a_2.

When a matrix **E** shown below is randomly generated from MVN(**0**, **V**) using the computer software, MATLAB (The MathWorks, Inc, 1999), random errors for the parameter estimates, a, b and c, are the diagonal elements of matrix **E**.

$$
\begin{array}{cc}
 & \begin{array}{ccc} \;\; d & \quad a_1 & \quad a_2 \end{array} \\
E = & \begin{bmatrix} .-.0731 & -.0129 & .0137 \\ -.0129 & .0285 & .0260 \\ .0137 & .0260 & .0590 \end{bmatrix}
\end{array}
$$

ERIC

The simulated (or observed) item estimates for the set of true parameters d=0.44, a_1=0.71 and a_2=0.53 are: d= 0.44+ (-0.0731), a_1=0.71+(0.0285)., and a_2=00.53+0.0590.

It should be noted that matrix E is randomly generated from the MVN(0, V) so that the values of its elements vary across replications. Therefore, the simulated item estimates for the set of true parameters d=0.44, a_1=0.71 and a_2=0.53 will be changed, along with the changes of the error matrix E. Theoretically, when a large number of replications is conducted, the standard deviations of the simulated item estimates, d, a_1 and a_2, will be close to the expected SEs of parameters, d, a_1 and a_2. They are .055, .040 and .036.

The above procedures of modeling measurement errors of item estimates is much easier to employ for some item-linking studies. In a research example conducted by Li and Lissitz (2000a) involving the existence of several multidimensional IRT item-linking methods, they attempted to examine which MIRT item-linking method is relatively less sensitive to the random (or sampling) errors of item parameter estimates. The above procedures for modeling random errors can be incorporated in the following procedures for this type of study.

1.Create the base test: Choose a set of item parameters for the base test. We treat these item parameters as known parameters.

2. Create the linked test: Assume item linking coefficients are known and generate a set of item parameters for the linked test by using these known linking coefficients.

3.1. Model random errors for the base-test item parameters: Each simulated item estimate from a set of parameters of an item is computed by summing the expected random error and the corresponding known (or true) item parameter. Expected random error was generated as a random value using the method outlined above.

3.2. Model measurement errors for the linked-test item parameters: Use the same method outlined in Step 3.1.

4. Estimate the equating coefficients: Estimate the equating coefficients based on two sets (base and linked) of item parameter estimates.

(5) Replication: Repeat Steps 3.1. 3.2 and 4 many times, which results in a large number of estimates for each individual item-linking coefficient; and calculate the BIAS (average difference between estimated and true values and RMSE (root mean squared error) of the item-linking coefficient estimate.

It should be noted that Step 3 is an alternative method of predicting the random errors of item estimates. Comparing the analytical approach with the replication approach to modeling measurement errors of item estimates, the analytical approach will save an enormous amount of time and energy in test data generation and item calibration for some types of research.

Using the analytical approach for modeling random errors of item estimates has its theoretical limitations. As indicated, measurement errors of item estimates are assumed to be distributed as **MVN** ($\underline{S}$, **V**). The analytical approach is used to model the "units of measurement errors for item estimates" (known as SEs of item estimates, associated with the matrix **V**). In addition, modeling the "points of origin of measurement error for item estimates" (known as the BIAS of item estimates, indicated by the vector $\underline{S}$) is another key issue to be considered. Although we might assume $\underline{S}$ to be $\underline{0}$, for simplicity, ML is a biased estimator (Anderson & Richardson, 1979) and the degree of bias depends upon the sample size. This issue of modeling $\underline{S}$ needs to be further explored in the future for better prediction of measurement errors of item

ERIC

estimates. If the bias of item estimates has a strong effect on the research being investigated, the analytical approach to modeling measurement error may not be appropriate.

V. Summary and Conclusions

The empirically determined SEs of MIRT MMLE/Bayesian item estimates for the forty items (ACT-Form 24b, 1985) were calculated for gauging the feasibility of utilizing MIRT models in real testing programs. Their magnitudes of SEs were all less than .2 and seemed to be reasonably small. This empirical finding connotes that MIRT 's item estimates can be reasonably stable so that the use of MIRT models will gradually become accessible in practice.

The analytical SEs of MIRT ML item estimates for those forty items were also calculated and compared with those empirically determined ones. The SEs of MIRT item estimates, in general, were quite similar when the two approaches were employed. This empirical comparison indicated that the analytical SEs of MIRT can approximately estimate the magnitudes of SEs for the MMLE/Bayesian item estimates. Accordingly, using the analytical approach to approximate SEs for the given item estimates and then tabulating them by the different combinations of item parameters (e.g., low d, low a_1 and high a_2) have been the subject of this paper. The tables associated with SEs of MIRT estimates provides a useful reference for test practitioners. Additionally, the 3-D diagrams that depicted the relationship between SEs for various item estimates demonstrate that extreme items (hard or easy) are more likely to have more measurement errors. In addition, an interesting issue was raised, namely, that the SE of the discrimination parameter is increasing as the discrimination parameters are

ERIC

getting larger. These findings deserve careful attention when using extreme items because they are more likely to be inaccurately calibrated.

The analytical approach can facilitate the simulation study of investigating which item-linking methods can better tolerate the random (or sampling) errors of item estimates. An example of how to utilize the analytical SEs of MIRT item estimates for this type of linking study was provided. Another application is that when researchers or test practitioners are interested in a set of item parameters that may be found in literature, in which the corresponding SEs of item estimates were not reported, the analytic approach provides them with a sense of how large standard errors of this set of item estimates might be under commonly-used situations.

Together the findings of this study support the use of MIRT models in practical applications as pointed out by (Miller, 1991) as well as the use of the analytical approach for approximating the SEs of MIRT MMLE/Bayesian item estimates when their SE estimates are practically unavailable or needed for simulation studies.

Further research is needed in several areas. As noted earlier, the SEs of MIRT item estimates evaluated in this study were based on a simpler two-dimensional MIRT model. The issue of whether the results found in this study can be generalized to more complex models (e.g., more than the two-dimensional MIRT models) needs to be explored. Also, with the sample size of 2000, we obtained quite a high degree of consistency of SE estimates between analytical and empirically determined approaches. Different sample sizes should be considered in future research for better understanding the effect of this factor on the consistency of SE estimates of item estimates as obtained between the two approaches. The findings obtained from this study were based on simulation test data that perfectly fit the presumed known model.

ERIC

This ideal data-model-fit condition can not be realized in real testing data and therefore similar research conducted by using real test data is necessary.

ERIC

References

Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. Journal of Educational Measurement, 29, 67-91.

Ackerman, T. A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. Applied Measurement in Education, 18, 255-278.

Anderson, J. A. & Richardson, S. C. (1979). Logistic discrimination and bias correction in maximum likelihood estimation. Technometrics, 21, 71-78.

Baker, F. B. (1992). Item Response Theory: Parameter Estimation Techniques. New York: Marcel Dekker, Inc.

Carlson, J. E. (1987). Multidimensional item response theory estimation [Computer software] (Research Report 87-19). Iowa City IA: ACT Inc.

De Ayala, R. J. & Sava-Bolesta, M. (1999). Item parameter recovery for the nominal response model. Applied Psychological Measurement, 23, 3-19.

Fraser, C & McDonald, R. P. (1988). NOHARM: Least Squares item factor analysis. Multivariate Behavioral Research, 23, 267-269.

Hambleton, R. K., Jones, R. W. & Rogers, H. J. (1993). Influence of item parameter estimation errors in test development. Journal of Educational Measurement, 30, 143-155.

Harwell, M. R., Stone, C. A., Hsu, T. , & Kirisci, L. (1996). Monte carlo studies in item response theory. Applied Psychological Measurement, 20, 101-125.

Li, Y. H. (1997). An evaluation of multidimensional IRT equating methods by assessing the accuracy of transforming parameters onto a target test metric. (Doctoral dissertation, University of Maryland, 1997). Dissertation Abstracts International, UMI Number 9816494.

Li, Y. H. & Lissitz, R. W. (2000a). An evaluation of the accuracy of multidimensional IRT linking. Applied Psychological Measurement, 24, 115-138.

Li, Y. H. & Lissitz, R. W. (2000b, April). Applications of the analytically derived asymptotic standard errors of IRT item parameter estimates. Paper presented at the annual meeting of the American Educational Research Association. New Orleans, LA.

Miller, T. R. (1991). Empirical estimation of standard errors of compensatory MIRT model parameters obtained from the NOHARM estimation program. (Research report 91-2). Iowa City, IA: The American College Testing.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. Applied Psychological Measurement, 16, 159-176.

Reckase, M. D. (1985). The difficulty of items that measure more than one ability. Applied Psychological Measurement, 9, 401-412.

Reckase, M. D. & Hirsch, T. M. (1991, April). Interpretation of number correct scores when the true number of dimensions assessed by a test is greater than two. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.

Stocking, M. L. (1990). Specifying optimum examinees for item parameter estimation in item response theory. Psychometrika, 3, 461-475.

The MathWorks, Inc. (1999). MATLAB (Version 5.3): The language of technical computing [Computer program]. Natick MA: The MathWorks, Inc.

Thissen, D. & Wainer, H. (1982). Some standard errors in item response theory.

Psychometrika, 47, 397-412.

Wilson, D., Wood, R. & Gibbons, R. D. (1991). TESTFACT: Test scoring, item statistics, and item factor analysis. Mooresville, IN: Scientific Software.

ERIC Full Text Provided by ERIC

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

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE
(Specific Document)

TM033731

I. DOCUMENT IDENTIFICATION:

Title: The Consistency between the Empirical and the Analytical Standard Errors of Multidimensional IRT Item Estimates

Author(s): Yuan H. Li   Yu N. Yang

Corporate Source:

Publication Date:

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 ☐
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 then ERIC employees end its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries end other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: Yuan H Li

Printed Name/Position/Title: Yuan H. Li   Statistical Specialist

Prince George's County Public Schools,
Department of Test Development and Admin.
Room 205,
14201 School Lane,
Upper Marlboro, MD. 20772

Telephone: 301-952-6764   FAX: 301-952-6228
E-Mail Address: jeffli@pgcps.org   Date: 3/10/02

(over)